High-resolution Analysis of a QTL for Resistance to Stagonospora Nodorum Glume Blotch on Chromosome 3B of Bread Wheat (*Triticum aestivum* L.)

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Margarita Shatalina

aus

Russland

Promotionskomitee

Prof. Dr. Beat Keller (Vorsitz und Leitung der Dissertation)

PD Dr. Thomas Wicker

Prof. Dr. Bruce McDonald

Zürich, 2013

Memory of Margarita Skurikhina and Mark Kotlyar

TABLE OF CONTENTS

SUMMARY

Stagonospora nodorum is a necrotrophic fungal pathogen causing Stagonospora nodorum glume (SNG) and leaf blotch (SNL) in bread wheat. Both diseases regularly cause severe yield losses in many wheat-growing areas. Previous QTL studies have shown that SNG and SNL resistance have a different genetic basis and that the major QTL for SNG resistance in the 'Arina' x 'Forno' population is located on the short arm of chromosome 3B. The first aim of the thesis was to map this QTL for SNG resistance (*QSng.sfr-3BS*) with a high resolution. For construction of the genetic map we were able to combine the information from the newly developed physical map of chromosome 3B and the SNP and InDel markers which were identified from our analysis of sequences of the flow-sorted chromosomes 3B of 'Arina' and 'Forno'. Surprisingly, artificial field infection tests of the fine-mapping population developed from a cross of susceptible cultivar 'Forno' and resistant cultivar 'Arina' revealed a co-segregation of SNG resistance and plant height. Both traits are highly correlated and a single marker analysis of the phenotypic data detected that SNG resistance and plant height are associated with the same small genetic region. A genetic co-localization of plant height and SNG resistance suggests that they are most likely controlled by the same gene(s) from the detected region. The analysis also revealed a second region linked to SNG resistance which is not linked to plant height. This hypothesis of two loci conferring SNG resistance on chromosome 3B indicates the complexity of the mechanism for race non-specific SNG resistance underlying the target QTL.

Bread wheat (*Triticum aestivum* L.) has a large repetitive genome and no reference sequence is currently available. Therefore, the development of molecular markers, essential for genetic mapping, is challenging. At the moment, the main strategy to approach the wheat genome analysis is to reduce its complexity. This strategy is being applied by flow sorting of individual chromosomes. The second aim of the thesis was the high-throughput development of genetic markers based on the sequences of flow-sorted chromosomes. We used next-generation sequencing (Illumina technology) to sequence isolated chromosomes of cultivars 'Arina' and 'Forno'. Assembled short contigs allowed us to eliminate the repetitive regions from analysis and to identify putative genes using the genome sequences of a model plant for grasses – *Brachypodium distachyon,* which has syntenic relationships with wheat. We focused on the identification of SNPs between two cultivars within the contigs containing putative genes in order to increase the amplification specificity. Additionally, the positions of the coding sequences in the *B. distachyon* genome were used to predict the approximate position of newly developed marker on chromosome

3B of wheat. Selected potential SNP-markers were mapped using KASPar technology resulting in a SNP-map for the entire chromosome 3B of the 'Arina' x 'Forno' population. Interestingly, this strategy of SNPs selection resulted in a high percentage of successfully mapped markers, which is unusual for SNP mapping in hexaploid wheat. The success rate shows a high potential for application of precise chromosome-specific SNP identification in genetic mapping. This is particularly relevant if the cultivars of interest have a similar genetic background and might not be efficiently genotyped by common SNP arrays.

ZUSAMMENFASSUNG

Stagonospora nodorum ist ein nekrophiler, pathogener Pilz, der Blatt- und Spelzenbräune in Brotweizen verursacht. Beide Krankheiten führen in der Regel zu einschneidenden Ernteausfällen in vielen Weizen-Anbaugebieten. Vorhergehende QTL Studien haben gezeigt, dass Blatt- und Spelzenbräune Resistenzen von verschiedenen genetischen Loci herrühren, und dass der wichtigste QTL für Spelzenbräune Resistenz in der 'Arina' x 'Forno' Population auf dem kurzen Arm des Chromosomes 3B zu finden ist. Das zweite Ziel dieser Dissertation war es, diesen QTL für Spelzenbräune Resistenz (*QSng.sfr-3BS*) mit einer hohen Auflösung zu kartieren. Um eine genetische Karte zu konstruieren, waren wir in der Lage, die Informationen aus der kürzlich entwickelten physischen Karte des Chromosomes 3B mit den von uns gefundenen SNP und InDel Markern zu verbinden. Überraschenderweise zeigten Infektionstests an der Feinkartierungs-Population aus der Kreuzung zwischen dem anfälligen Kultivar 'Forno' und dem resistenten Kultivar 'Arina' im Feldversuch, dass eine Segretation nicht nur bezüglich der Spelzenbräune Resistenz sondern auch bezüglich der Höhe der Pflanzen zwischen den nahe isogenen Linien zu beobachten war. Beide Eigenschaften korrelierten stark und eine Einzelmarker-Analyse kombiniert mit den Phänotyp-Daten zeigte darüber hinaus auf, dass sowohl Spelzenbräune Resistenz wie auch die Pflanzenhöhe mit der selben, kleinen genetischen Region assozieren. Eine solche genetische Co-Lokalisation der beiden Eigenschaften deutet darauf hin, dass beide wahrscheinlich von/vom den/m gleichen Gen(en) kontrolliert werden. Die Analyse zeigte zudem, dass eine zweite Region, die zwar mit Spelzenbräune Resistenz jedoch nicht mit der Pflanzenhöhe gekoppelt ist, zu den beobachteten Unterschieden führt. Die Hypothese, wonach zwei Loci auf Chromosom 3B nötig für die rassen-unspezifische Spelzenbräune Resistenz sind, weist auf die die Komplexität des Mechanismus dieses QTLs hin.

Brotweizen (*Triticum aestivum L.*) besitzt ein grosses, repetitives Genom, für das derzeit noch keine Referenz Sequenz erhältlich ist. Die Entwicklung von molekularen Markern, eine essentielle Voraussetzung für die genetische Kartierung, stellt daher eine grosse Herausforderung dar. Die derzeit hauptsächlich angewendete Strategie besteht darin, die Komplexität des Genoms zu reduzieren. Dabei werden in einem ersten Schritt die einzelnen Chromosomen mittels “fluss sortieren” aufgetrennt. Eines der Ziele dieser Dissertation war die Entwicklung von genetischen Markern für Chromosom 3B, basierend auf einer Kreuzung der Sorten 'Arina' x 'Forno' und der Sequenzierung dieser Elternlinien. Dazu wurden Sequenzier-Technologien der nächsten Generation (Illumina) verwendet. Durch das Assemblieren der Sequenzen wurden grosse repetitive Regionen herausgefiltert.

Auf den übrigen Contigs konnten dann, anhand der nahen, syntenischen Verwandschaft zur Modell-Pflanze *Brachypodium distachyon*, potentielle Gen-Kandidaten identifiziert werden. Bei der Suche nach SNPs zwischen den beiden Sorten haben wir uns auf diese Regionen beschränkt, um eine möglichst hohe Spezifität zu erzielen. Zusätzlich konnten die Positionen der Gene im Genom von *B. distachyon* benutzt werden, um die ungefähre Position der neu entwickelten Marker auf Chromosom 3B in Weizen vorher zu sagen. Die nach diesen Kriterien ausgewählten SNPs wurden dann mit Hilfe der KASPar-Technologie kartiert, was zu einer SNP-Karte für das gesamte Chromosom 3B der 'Arina' x 'Forno' Population führte. Die durch dieses Verfahren erzielte SNP Auswahl führt zu einem sehr hohen Anteil an erfolgreich kartierten Markern, was für das SNP Kartieren in hexaploidem Weizen unüblich ist. Diese vielversprechende Erfolgsquote zeigt das grosse Potential dieser Strategie für die präzise, chromosomen-spezifische Identifikation von SNPs für die genetische Kartierung. Diese Technik ist besonders interessant, wenn die untersuchten Sorten einen ähnlichen genetischen Hintergrund besitzen und übliche Strategien nicht effizient sind.

CHAPTER 1

General introduction

Domestication of bread wheat species and their agricultural importance

Wheat domestication occured in the Near-East, in an area called the Fertile Crescent near the upper reaches of the rivers Tigris and Euphrates (Zohary and Hopf 2001; Salamini *et al.* 2002). This region nowadays corresponds to south-eastern Turkey, northern Syria, Israel, Jordan and Iraq. The domestication of einkorn (*Triticum monococcum* L.), emmer wheat (*Triticum dicoccum* L.), barley (*Hordeum vulgare* L.) and rye (*Secale cereale* L.) occured between 10,500 and 7,800 years ago (Charmet 2011). Since early agricultural practices did not include direct selection for crop domestication, domesticated and wild forms were grown side-by-side during a few centuries. There is a typical hallmark of domestication – nonshattering seed. Spikes possessing this hallmark were found together with the wild forms in archaeological sites (Balter, 2007). Hexaploid bread wheat originated approximately 9,000 years ago from a hybridization event between the allotetraploid domesticated emmer wheat (*Triticum turgidum* spp. *dicoccoides,* 2n=4x=AABB) and the diploid wild goatgrass *Aegilops tauschii* L. (2n=2x=DD). The donor of the A genome in bread wheat is *Triticum urartu. Aegilops speltoides* L. (SS) has been suggested as the closest species to the B-genome contributor (Sarkar and Stebbins 1956; Kilian 2007). Wild allotetraploid wheat (*T. turgidum,* genome AABB, 2n=4x=28) was formed approximately 0.5 million years ago (Huang *et al.* 2002).

Nowadays, hexaploid bread wheat belongs to the three most cultivated crops in the world. In some regions wheat historically is the most important source of calories for the human diet. According to the FAO, 682.5 million tons of wheat were harvested in 2011 (www.faostat.org). Both hexaploid bread wheat (*Triticum aestivum* L., 2n=6x=42, AABBDD) and tetraploid durum wheat (*Triticum durum* L., 2n=4x=28, AABB), which is used for production of pasta and low-rising bread, account for more than 20% of human food calories and are a staple food for 40% of the world's population (Peng *et al.* 2011).

Wheat is grown on 17% of all crop growing areas, in variable climatic conditions: from northern altitudes of 67ºN (Norway, Finland, Russia) to 45ºS (Argentina, Chile). The main wheat-growing areas are situated in temperate and southern Russia, central regions of the US, southern Canada, Mediterranean area, northern China, India, Argentina and Australia (www.faostat.org).

As the world population is constantly increasing, the importance of food security is increasing as well. Therefore, the improvement of the wheat yield potential and the reduction of yield losses are needed. One of the critical factors influencing yield are disease epidemics. Global yield loss due to all wheat diseases was estimated to be 12.4% annually (Oerke *et al.* 1994). Fungicides have a wide application for crop protection against diseases. However, this method of disease management has several disadvantages. First, fungicides often are unaffordable for the farmers in developing countries and second, the emergence of fungicide-resistant pathogen races is a growing threat (Duveiller *et al.* 2007). An alternative method for disease control is to screen the germplasm for disease resistance genes from resistant cultivars, landraces or wild wheat relatives and to incorporate them into cultivated wheat.

Disease resistance in plants

The different types of disease resistance found in plants can be characterized by their genetic control, the spectrum of specificity and the durability. Disease resistance can be either under monogenic or polygenic control. Race-specific genes confer disease resistance only to a particular pathogen race, whereas race non-specific genes are effective against many pathogen races (so-called broad-spectrum resistance). Another important aspect for wheat breeding is the classification of disease resistance as durable and non-durable. Johnson (1983) defined durable resistance as “resistance that remains effective while a cultivar possessing it is widely cultivated in an environment favoring the disease”.

Monogenic resistance is often race-specific and non-durable. Products of race-specific resistance genes (so-called *R* genes) directly or indirectly interact with a corresponding avirulence gene (*Avr* gene) product of the pathogen which is also called effector. Fungal effectors (e.g. AVR proteins) are defined as any small molecules (mainly proteins) inducing or suppressing natural plant immune responses. Upon recognition of this specific avirulence (*Avr*) protein by the *R* protein, a resistance response against this specific pathogen is triggered. Resistance in this case is mostly based on the hypersensitive reaction (HR) – local cell death at the site of attempted pathogen invasion. This “gene-for-gene” interaction was first described by Flor (1955). *R* genes typically encode NBS-LRR type of proteins (nucleotide binding and leucine-rich repeat domains). An example of a NBS-LRR resistance gene is the cloned race-specific *Pm3* gene (Yahiaoui *et al.* 2004) that

provides resistance to some isolates of *Blumeria graminis* f. sp. *tritici,* a fungus causing wheat powdery mildew. Pathogens avoid recognition by *R* genes by eliminating or modifying effectors. Because such loss-of-function mutations occur frequently, *R* gene based resistance is often short-lived in the field. For instance, an aggressive stem rust *(Puccinia graminis* f. sp. *tritici)* race Ug99 (first emerged in Uganda in 1998) carries virulence to several wheat stem rust resistance genes. One of those genes is *Sr31,* a resistance gene that has been translocated from rye (*Secale cereale* L.) to wheat, and has been very effective against wheat stem rust. To date, Ug99 lineage has spread to eastern Africa, Zimbabwe, South Africa, Sudan, Yemen, and Iran. It remains a major threat to wheat grown worldwide, because 90% of the varieties are susceptible to Ug99 (reviewed in Singh *et al.* 2011). However, there are also examples of monogenic, resistance genes that are not race-specific. For example, the *mlo* gene in barley (*Hordeum vulgare* L.) controls a broad spectrum resistance to barley powdery mildew (*Blumeria graminis* f. sp. *hordei*) (Buschges *et al.* 1997). Another race non-specific resistance gene is *Sr33,* which confers broad spectrum resistance to stem rust in wheat (Mago *et al.* 2002).

Polygenic resistance is often race non-specific and durable. Polygenic resistance provides protection against many pathogen races and has been found to be efficient for a longer period of time (Parlevliet 2002; Lindhout 2002). To overcome polygenic resistance fungal pathogens need to undergo multiple gain-of-function mutations, which occur rarely. Additionally, if polygenic resistance is based on a general physiological process leading to disease escape, pathogens would need also to develop a new lifestyle feature to overcome the resistance. Thus, for effective disease control, the incorporation of broad-spectrum, durable resistance in wheat cultivars is preferred. Polygenic resistance is usually identified using the statistical concept of quantitative trait loci (QTL). QTL mapping detects genomic regions which are associated with the observed resistance phenotype. The leaf rust resistance gene *Lr34* and the stripe rust resistance gene *Yr36* in wheat are examples of disease resistance genes identified by mapping of major QTLs (Krattinger *et al.* 2009; Fu *et al.* 2009).

Stagonospora nodorum – a nectrotrophic fungal pathogen of wheat[1]

Phaeosphaeria nodorum (anamorph – asexual form – *Stagonospora nodorum)* is a necrotrophic fungus which infects and kills wheat leaf and glume tissue and feeds on the organic compounds of the dead cells during its life cycle (Figure 1). Stagonospora nodorum blotch (SNB) affects wheat grown under humid conditions and mild

1 – this text is a part of the book chapter "Identification and implementation of resistance: genomics-assisted use of genetic resources for breeding against powdery mildew and Stagonospora nodorum blotch in wheat" by L. Selter, M. Shatalina*, J. Singla* and B. Keller, in preparation*

**authors contributed equally*

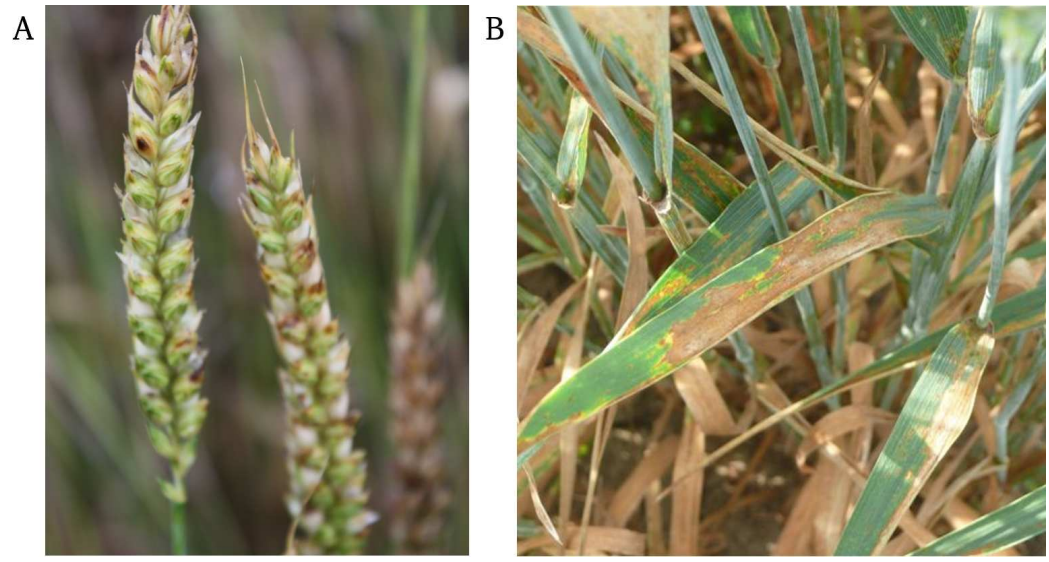

Figure 1. Symptoms of wheat Stagonospora nodorum glume (A) and leaf blotch (B). SNB manifests itself as brown necrotic spots on the spikes and leaves of wheat

temperatures in Europe, South America, Central Asia and North Africa. A wide range of fungicides are efficiently applied in the areas with SNB infections. Reports about fungicide-resistant isolates of *S. nodorum* are very rare. However, the possibility of their emergence remains a threat in regions with extensive fungicide application (Oliver *et al.* 2012).

Disease epidemics of SNB depend mostly on three factors: prevalence of inoculum, the genetic constitution of grown cultivars, and to a large extent on environmental conditions (Duveiller *et al.* 2007).

Susceptibility to Stagonospora nodorum leaf blotch is mediated by fungal host-selective toxins[1]

To kill wheat cells, *S. nodorum* produces host-selective toxins (also called necrothrophic effectors). These host-selective toxins (HST) interact with the plant host in a mirrored gene-for-gene interaction. According to the classical gene-for-gene model developed by Flor (1955), a pathogen is only able to invade the host successfully if the pathogens virulence factor is not recognized by the corresponding plant *R* gene. In the mirrored gene-for-gene interaction of Stagonospora nodorum leaf blotch (SNL), an infection will be successful only if the HST is recognized by a corresponding susceptibility gene of the plant (Figure 2; Friesen *et al.* 2007).

This type of interaction was identified as the cause of a few fungal diseases in different plant species (Table 1; Mengiste, 2012). All three cloned plant toxin-sensitivity genes had a NBS-LRR structure, which resembled a typical plant resistance gene.

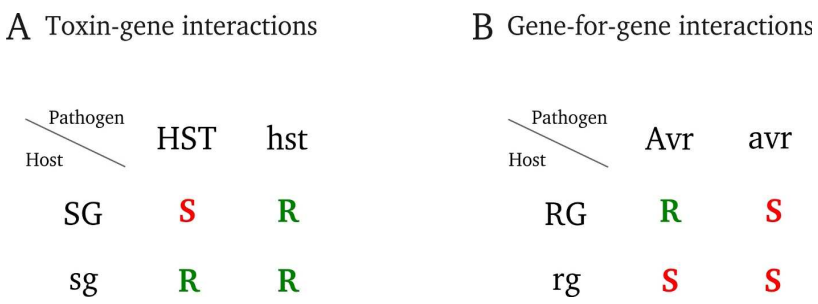

Figure 2. Toxin-gene and gene-for-gene interactions in wheat

(A) Toxin-gene interactions: only the interaction of host-selective toxin (HST) from the fungus and corresponding sensitivity gene (SG) of wheat results in susceptibility of the plant; (B) Gene-for-gene interactions: Avirulence protein (Avr) of the fungus triggers resistance response only if corresponding resistance gene (RG) is present in wheat.

Table 1. Cloned plant toxin-sensitivity genes which interact with fungal toxins resulting in susceptible disease response

Plant species	Fungal pathogen	Toxin	Susceptibility gene	References
Sorghum *(Sorghum bicolor)*	*Periconia circinata*	PC toxin	*Pc* (NBS-LRR)	Nagy *et al.* 2007
Arabidopsis thaliana	*Cochliobolus victoriae*	victorin	*LOV1* (NBS-LRR)	Lorang *et al.* 2007
Wheat *(Triticum aestivum)*	*Stagonospora nodorum*	ToxA	*Tsn1* (NBS-LRR)	Faris *et al.* 2010

The recently sequenced genome of *S. nodorum* provided the opportunity to study the genetic basis of pathogenicity together with other features of the fungal lifestyle. The genome size was estimated to be 37.2 Mbp (Hane *et al.* 2007) and gene predictions and expressed sequence tags (EST) library analysis suggested that the genome contains at least 10,762 genes. Interestingly, a large number of identified genes were predicted to encode secreted proteins with no similarity to any known genes. Possibly, new HST genes are among these genes. For instance, the HST SnTox1 was identified by screening the whole *S. nodorum* genome for suitable candidates and then testing them in infection experiments (Liu *et al.* 2012).

Different strains of *S. nodorum* produce a range of HST. Five different toxins SnToxA, SnTox1, SnTox2, SnTox3 and SnTox4 have been identified until now (Liu *et al.* 2004; Friesen *et al.* 2006; Friesen *et al.* 2007; Abeysekara *et al.* 2009). The susceptibility genes for all five toxins were mapped to different regions of the wheat genome: *Tsn1* interacts with ToxA and this interaction accounts for 58% of the phenotypic variation observed in

the ITMI population derived from a cross of synthetic hexaploid wheat 'W-7984' and cultivar 'Opata' (Liu *et al.* 2004). The interaction of *Snn2-SnTox2* explains 47% of the phenotypic variation and *Snn3-SnTox3* accounts for 17% of the variation in the population of hard red spring wheat line 'BR34' and cultivar 'Crandin'. Interaction of Snn4-SnTox4 identified in the RIL population of 'Arina' and 'Forno' explains 41% of the phenotypic variation (Abeysekara *et al.* 2009).

Interestingly, each fungal toxin-wheat gene interaction is qualitative, but they contribute to the resistance response in a quantitative manner. For example, SnToxA-*Tsn1* and SnTox2-*Snn2* have additive effects during the infection (Oliver *et al.* 2012). In other words, the susceptibility genes have additive effects if multiple compatible interactions are acting at the same time. Therefore, the resistance to Stagonospora nodorum leaf blotch is quantitatively inherited (Abeysekara *et al.* 2009). *Tsn1* confers sensitivity to SnToxA and is located on the long arm of chromosome 5B. It was recently cloned using a classical chromosome walking approach (Faris *et al.* 2010). Functional validation revealed that *Tsn1* has a resistance gene-like structure consisting of a nucleotide-binding, leucine-rich repeat (NBS-LRR) and a serine/threonine protein kinase (S/TPK) domain. Faris *et al.* (2010) suggested that in the case of the Tsn1-ToxA interaction, *S. nodorum* may have subverted a wheat defence mechanism, acquired by plants to protect them from other pathogens.

SNB resistance on wheat glumes and leaves are genetically independent[1]

Genetic studies suggest that resistance to SNB is complex and in most cases polygenic (Table 2; Bostwick *et al.* 1993; Du *et al.* 1999; Fried and Meister 1987; Scott *et al.* 1982). However, monogenic resistance was also identified in some wheat varieties (Kleijer *et al.* 1977; Ma and Hughes 1995; Murphy *et al.* 2000). The resistance responses to SNB on leaves and glumes are genetically independent (Francki *et al.* 2011). Several QTL controlling partial resistance to Stagonospora nodorum leaf blotch in seedlings were identified on chromosomes 2B, 3B, 5B and 5D using a double haploid population derived from a cross of winter wheat cultivars 'Liwilla' and 'Begra' (Czembor *et al.* 2003). However, their effect on adult plants was not tested. For example, Francki *et al.* (2011) discovered three QTLs using a cross of winter wheat cultivars 'P92201D5' x 'P91193D1' and spring wheat 'EGA Blanko' x 'Millewa'. Two of them, located on chromosomes 1BS and 2AS respectively, did not correlate with any known toxin sensitivity genes. In contrast, the third QTL on chromosome 5BL was associated with *Tsn1*-ToxA insensitivity.

Several QTLs for SNB resistance on wheat glumes were identified. These QTLs were assigned to the chromosomal locations that differ from leaf SNB resistance QTLs. For instance, two major Stagonospora glume blotch QTLs (explaining up to 20 and 30% of

phenotypic variation) were identified in the RIL population of cultivars 'Arina' and 'Forno' and located on chromosomes 3B and 4B (Schnurbusch *et al.* 2003). Two other QTLs detected in a RIL population of lines 'P92201D5' x 'P91193D1' were assigned to chromosome 2D (Uphaus *et al.* 2007). Another Stagonospora glume blotch resistance QTL was identified in double haploid population 'WAWHT2074' x '6HRWSN125' on chromosome 4B (Shankar *et al.* 2008). Independent genetic control of resistance to SNB in glumes and leaves combined with diverse resistance on different stages of plant growth suggests that the best strategy for breeding is to combine the different genetic loci and take advantage of their additive effects.

Table 2. Summary of identified QTLs for resistance to Stagonospora glume and leaf blotch in bread wheat. Modified from Francki *et al.* 2013

Plant tissue	Population	Chr	Total variation	Toxin-based	Reference
Seedling	Liwilla x Begra, DH	2B	16%		Czembor *et al.* 2003
		5B	30%		
		5D	37%		
Seedling	W7984 x Opata85	1B	27-58%	yes (*Snn1*)	Liu *et al.* 2004
		4B	6-9%	yes	
Seedling	Alba x Begra, DH	6A	36%		Arseniuk *et al.* 2004
Seedling	BR34 x Grandin, RIL	2DS	24%	yes (*Snn2*)	Friesen *et al.* 2009
		5AL	11%	yes	
		5BL	37%	yes (*Tsn1*)	
Flag leaf	Forno x Oberkulmer, RIL	2D	20%		Aguilar *et al.* 2005
		4B	17%		
		7B	12%		
Flag leaf	WAWHT2074 x 6HRWSN125, DH	2D	8-17%		Shankar *et al.* 2008
Flag leaf	BR34 x Grandin, RIL	1BS	10%	yes	Friesen *et al.* 2009
		2DS	12-15%	yes (*Snn2*)	
		5AL	12-18%	yes	
		5BL	11-18%	yes (*Tsn1*)	
Flag leaf	P92201D5 x P91193D1, RIL	2A	11-21%		Francki *et al.* 2011
Flag leaf	EGA Blanco x Millewa, DH	1B	15-16%		Francki *et al.* 2011
		5B	8-16%		
Glume	Arina x Forno, RIL	3BS	12-30%		Schnurbusch *et al.* 2003
		4BL	7-22%		
Glume	P92201D5 x P91193D1, RIL	2DL.1	12-38%		Uphaus *et al.* 2007
		2DL.2	5-6%		
Glume	WAWHT2074 x 6HRWSN125, DH	4B	8-19%		Shankar *et al.* 2008

Breeding for resistance to Stagonospora nodorum blotch[1]

Based on the recent findings on HST in the *S. nodorum* – wheat pathosystem, it is evident that the presence or absence of specific toxin receptors in the widely grown wheat cultivars will have a significant impact on disease prevalence. Recently it was shown that there are significant differences between the frequencies of toxin presence in *S. nodorum* isolates originating from different geographical regions (McDonald *et al.* 2013). This suggests that the presence/absence of sensitivity genes in the cultivars grown in particular regions has a strong effect: whenever a cultivar contains the sensitivity gene corresponding to a specific HST, the presence of this toxin will be of selective advantage for the pathogen and races with the toxin will increase in frequency. On the other hand, if the sensitivity gene is absent, there will be no selective advantage for having the effector and it is likely that the frequency of such races will decrease.

These findings immediately suggest that a breeding strategy which has the goal to eliminate from the germplasm as many of the relevant susceptibility genes as possible might be effective (it remains to be determined which ones belong to this group in addition to *Tsn1*). This has not yet been tried before but has considerable potential to reduce the problem of SNB based on diagnostic markers for a limited subset of susceptibility genes. The markers would allow the elimination of all breeding material with active susceptibility genes. At this stage, only the Tsn1 receptor is cloned and more map-based cloning projects are needed to molecularly isolate the other toxin sensitivity genes. Ideally such an effort to eliminate susceptible lines would be coordinated in large geographical areas to ensure success and reduce the frequency of toxin sensitivity genes. Such a project is ongoing in Australia to eliminate the *Tsn1* gene from commercial germplasm (Zhang *et al.* 2009; Oliver and Solomon 2010; Waters *et al.* 2011). Elimination of toxin sensitivity genes should confer durable resistance. In contrast to *R*-gene based resistance, where a loss of function of *Avr* would lead to susceptibility, a gain of function for HST would be needed for susceptibility if wheat sensitivity gene is eliminated. However, the possible consequences of toxin sensitivity genes elimination are not yet clear, as they might play a role in resistance to other pathogens.

Based on the molecular advancements in understanding the *S. nodorum* – wheat pathosystem, future resistance breeding efforts will possibly rely more on molecular markers for selecting against susceptibility (receptor) genes and not only depend on phenotyping under field conditions. It will be interesting to see if similar types of genes are responsible for resistance to Stagonospora nodorum glume blotch. As resistance in the glume is inherited independently from resistance in the leaf, other genetic factors must be involved (Schnurbusch *et al.* 2003).

Synteny among grasses and the use of model species for mapping

Bread wheat (*T. aestivum*) is an allohexaploid plant with a large and highly repetitive genome. Its size estimate is approximately 17 Gb and it contains up to 80% of repetitive elements (Bennett and Smith 1976; Hollister and Gaut 2009). To date, there is no reference sequence available for the genome of bread wheat. Therefore, gene mapping and marker development are hindered by the enormous genome size and absence of the reference sequence. The sequence composition of the three sub-genomes is very similar, but not identical (Gu *et al.* 2004). The similarity of the three sub-genomes of bread wheat is a major problem for the identification of sub-genome-specific polymorphisms. Repetitive genome content and the similarity of sub-genomes are increasing the risk of unspecific amplification of genetic markers. Additionally, target traits (e.g. *R*-genes) are often found in only few genotypes, particularly in elite cultivars. These elite cultivars have very low number of polymorphisms for historical reasons: modern wheat varieties were bred from a limited number of landraces which resulted in a diversity bottleneck. As a consequence, modern wheat varieties are approximately 99.9% similar on the genome level (Ravel *et al.* 2006; Chao *et al.* 2009, Trick *et al.* 2012).

Among the genomes of grasses (*Poaceae*) a high conservation of the gene order is very typical due to their recent common origin 50-75 MYA. The genomes of *Brachypodium distachyon* L. (2n=2x=10, 0.272 Gb), rice (*Oryza sativa* L., 2n=2x=24, 0.4 Gb), maize (*Zea mays* L., 2n=2x=20, 2.5 Gb), sorghum (*Sorghum bicolor* L., 2n=2x=20, 0.73 Gb), barley (*Hordeum vulgare* L., 2n=2x=14, 5.1 Gb) and wheat (*Triticum aestivum* L., 2n=6x=42, 17 Gb) share syntenic sets of genes (Gale and Devos, 1998; Devos, 2005). Synteny describes the presence of chromosomal blocks with a conserved gene order among different genomes. The genomes of *B. distachyon*, rice, maize, sorghum and barley were sequenced and their gene content was annotated (International Rice Genome Sequencing Project 2005; Paterson *et al.* 2009; Schnable *et al.* 2009; International Brachypodium Initiative 2010; International barley sequencing consortium 2012). Syntenic relationships between bread wheat and sequenced grasses are very useful for genetic mapping in wheat. Synteny allows to “unlock” the genome without a reference sequence by building a possible gene order. This has been done for chromosomes of barley and wheat, where a so called “genome zipper” has been created (Mayer *et al.* 2011; Hernandez *et al.* 2012). Additionally, synteny allows to characterize a target region in respect to possible number of genes, order of genetic markers and physical size in syntenic species.

However, breakdowns of synteny also have been described. There are synteny perturbations between homeologous chromosomes of wheat caused by locus duplications and deletions (Akhunov *et al.* 2003). In has been shown that in chromosome 3B synteny was disrupted by nonsyntenic genes which were interspersed along the syntenic backbone

(Choulet *et al.* 2010). Therefore, the syntenic “genome zipper” is a very useful tool, but for high-resolution mapping it might be not sufficient due to the presence of nonsyntenic genes or regions.

Genome sequencing of bread wheat

The wheat genome is currently sequenced by the International Wheat Genome Sequencing Consortium (IWGSC) and the task will be completed in a few years (IWGSC, www.wheatgenome.org). Recently, the whole-genome shotgun sequence of cultivar 'Chinese Spring' become available (Lai *et al.* 2012; Brenchley *et al.* 2012). This shotgun sequence cannot be called a “complete” genome sequence of wheat due to the low read coverage and, consequently, only partial genome assembly. However, 454 pyrosequencing produced relatively long reads and this allowed Brenchley *et al.* (2012) to assemble gene representations for 94000 – 96000 genes and assign them to A, B and D genomes using gene sequences of *T. monococcum, Ae. speltoides* and *Ae. tauschii,* respectively. This gene content information can be applied in evolutionary comparative genomic studies and in gene mapping projects as well.

Besides the shotgun sequencing, there are two other approaches for sequencing of the wheat genome. Both approaches reflect a general idea of reducing complexity in the large and repetitive wheat genome: (1) BAC-by-BAC sequencing of isolated chromosomes and (2) sequencing of diploid relatives. Sequencing of the diploid relatives was already successfully applied in other sequencing projects, for example, Shulaev *et al.* (2011) accessed the allo-octoploid genome of cultivated strawberry *(Fragaria* x *ananassa,* 2n=8x=56*)* via sequencing of the 240 Mb diploid genome of woodland strawberry (*Fragaria vesca,* 2n=2x=14). In wheat, the BAC-libraries of diploid wheat relatives – *Ae. tauschii*, *Ae. speltoides* and *T. urartu* were developed (Akhunov *et al.* 2005) to access A, B and D genomes of hexaploid wheat. Recently, both progenitors of A- and D-genomes were sequenced using shotgun sequencing approach. This sequencing approach resulted in a construction of the draft genome sequences for *A. tauschii* and *T. urartu* (Jia *et al.* 2013; Ling *et al.* 2013).

The BAC-by-BAC sequencing approach is the most time- and resource consuming, but also the most reliable as it includes a construction of physical maps of flow-sorted chromosomes (Paux *et al.* 2008; Lucas *et al.* 2013; reviewed in Dolezel *et al.* 2012) and sequencing of individual chromosomes. A minimal set of BAC clones covering the physical map, named minimal tiling path (MTP) will be identified for each chromosome. The sequencing of the MTP for each chromosome should provide a complete assembly including both genes and intergenic regions. Thus, the result of this strategy should be a chromosome-specific reference sequence of each chromosome (reviewed in Feuillet *et al.*

2012).

Recently, partial chromosome-specific survey sequences for all wheat chromosomes became available (Hernandez *et al.* 2012; IWGSC, www.wheatgenome.org). The sequences have different coverage depths from 1.5 fold to 80 fold and they can be accessed at IWGSC Survey Sequence repository at the Unité de Recherches en Génomique Info (http://wheat-urgi.versailles.inra.fr/Seq-Repository). Those sequences are not assembled in long scaffolds, nevertheless, they can be very useful for BLAST searches and to assign new markers or sequence fragments to individual chromosomes or sub-genomes. These survey sequences will also be used for creating a virtual gene order for each chromosome using comparative genomics and collinearity with other grasses (IWGSC, www.wheatgenome.org).

Molecular markers available for gene mapping in bread wheat

The essential condition for gene mapping is the availability of a sufficient number of genetic markers. The wheat research community made great efforts over the past twenty years to develop and share genetic markers through international mapping projects such as the International Triticeae Mapping Initiative (www.wheat.pw.usda.gov/ITMI).

Classical approaches to develop genetic markers are mainly very laborious and low throughput. Simple Sequence Repeats (SSR) belong to these classical markers. This type of markers is based on the length differences of microsatellite repeats between two cultivars. It is a sequence-based marker and can be designed using a small amount of sequence information. Another type of classical genetic markers is based on restriction fragment length polymorphisms (RFLP). Polymorphism detection is based on the difference in restriction sites between different wheat cultivars. This approach is labour-intensive, but requires no sequence information. Amplified fragment length polymorphisms (AFLP) is a PCR-based tool for detection of differences in localization of restriction sites. After digestion of genomic DNA, restriction fragments are amplified using adaptors and primers complementary to the sequence of the restriction sites, followed by detection of amplified fragments. DArT technology is also using restriction enzyme digestion of genomic DNA by rare and frequent cutters, followed by amplification and hybridization steps to detect differences between cultivars. This technology is high-throughput and not based on sequence information. These markers were extensively used in the past and still have many applications in various mapping projects.

With the increasing number of genomic wheat resources the type of utilized markers is shifting from non-sequenced based markers (RFLP, DArT, AFLP, SSR) to high-throughput sequence-based markers (ISBP, SNP) (reviewed in Paux *et al.* 2012). Insertion site based polymorphism (ISBP) markers are exploiting differences in the insertion sites of the

transposons. The number of ISBP markers and especially SNP (single nucleotide polymorphism) markers based on sequence information is constantly increasing with the increase of sequence information available for the genome of bread wheat. These markers can be genotyped in a high-throughput manner and are easily reproducible.

Perspectives of gene mapping in wheat

The increasing amount of available sequence information, the construction of "genome-zippers" and the growing number of high-throughput genetic markers have the potential to speed up genetic mapping in wheat.

In general, high-resolution mapping and cloning of genes in crops with a reference genome sequences is more straightforward. For example, genome resequencing allowed to conduct large SNP discovery projects in rice (McNally *et al.* 2009; Yamamoto *et al.* 2010) and in maize (Lai *et al.* 2010). The obtained sets of high throughput genetic markers serve as the basis for fast high-resolution mapping, large genome-wide mapping projects and haplotype studies. For example, high-throughput resequencing of 150 rice recombinant inbred lines by the Illumina technology identified 1 493 461 SNPs (Huang *et al.* 2009). These SNPs allowed to construct a genetic map by identifying recombination breakpoints. This allowed to localize a major QTL for plant height to a 100 kb region. In maize, the resequencing of six elite commercial maize lines with a 5.4-fold genome coverage lead to the identification of 1.2 million SNPs in non-repetitive regions (Lai *et al.* 2010). Recently, Huang *et al.* (2010) re-sequenced 517 rice landraces and identified ~3.6 million SNPs for constructing a haplotype map for genome-wide association study (GWAS) in rice, which was conducted for 14 agronomic traits in *Oryza sativa* spp. *indica*.

Thus, genetic markers developed by genome resequencing are high-throughput, very specific (can be developed for any given population) and time efficient. The availability of a reference sequence and, consequently, a large number of high-throughput molecular markers allows a faster detection of a small genetic interval linked to the trait of interest and identification of possible candidate genes.

Aim of the thesis

The aim of the thesis was the high-resolution mapping of a major QTL for resistance to Stagonospora nodorum glume blotch on chromosome 3B (*QSng.sfr-3BS*) of bread wheat. The first part of the thesis (Chapter 2) describes the genetic mapping of SNP markers in recombinant inbred lines of Arina x Forno as a "proof of concept". SNP markers were developed based on the comparative analysis of sequence assemblies from sorted chromosomes 3B for both cultivars. In the second part of the thesis (Chapter 3), we developed a population of near-isogenic lines, constructed a genetic map for the target interval and evaluated disease resistance of the mapping population in the field to narrow down the target region of QTL.

CHAPTER 2

Genotype-specific SNP map based on whole chromosome 3B sequence information from wheat cultivars Arina and Forno

Margarita Shatalina, Thomas Wicker, Jan P. Buchmann, Simone Oberhaensli, Hana Šimková, Jaroslav Doležel and Beat Keller

Published in Plant Biotechnology Journal (2013) 11: 23-32
doi: 10.1111/pbi.12003

Summary

Agronomically important traits are frequently controlled by rare, genotype-specific alleles. Such genes can only be mapped in a population derived from the donor genotype. This requires the development of a specific genetic map which is difficult in wheat because of the low level of polymorphism among elite cultivars. The absence of sufficient polymorphism, the complexity of the hexaploid wheat genome as well as the lack of complete sequence information make the construction of genetic maps with a high density of reproducible and polymorphic markers challenging. We developed a genotype-specific genetic map of chromosome 3B from winter wheat cultivars Arina and Forno. Chromosome 3B was isolated from the two cultivars and then sequenced to 10 fold coverage. This resulted in a Single Nucleotide Polymorphisms (SNP) database of the complete chromosome. Based on proposed synteny with the *Brachypodium* model genome and gene annotation, sequences close to coding regions were used for the development of 70 SNP-based markers. They were mapped on a Arina x Forno Recombinant Inbred Lines population and found to be spread over the complete chromosome 3B. While overall synteny was well maintained, numerous exceptions and inversions of syntenic gene order were identified. Additionally, we found that the majority of recombination events occurred in distal parts of chromosome 3B, particularly in hot-spot regions. Compared to the earlier map based on SSR and RFLP markers, the number of markers increased fourfold. The approach presented here allows fast development of genotype-specific polymorphic markers which can be used for mapping and marker-assisted selection.

Introduction

Bread wheat (*Triticum aestivum* L., 2n=6x=42), is one of the most important food crops in the world. The hexaploid genome of wheat has a size of approximately 17 Gb and consists of the three homoeologous genomes A, B and D. Therefore, most genes are present in three homoeologous copies. Additionally, the wheat genome is very repetitive and contains over 80% of repetitive elements (Bennett and Smith, 1976; Hollister and Gaut, 2009). The characterization of genes involved in agronomically important traits requires the identification of markers closely linked to the genes. Single Nucleotide Polymorphisms (SNPs) are one of the most reliable and reproducible types of sequence-based genetic markers. Sequence information for the hexaploid wheat genome has only recently begun to emerge and thus, the construction of genetic maps with a high density of genetic markers remains challenging.

Recently, the genomes of rice (*Oryza sativa* L.), sorghum (*Sorghum bicolor* L.) and *Brachypodium distachyon* were sequenced (International Rice Genome Sequencing Project, 2005; Paterson *et al.* 2009; International Brachypodium Initiative, 2010). Since these genomes have a relatively small size and the gene order in grasses is conserved among different members of the family (Gale and Devos, 1998; Devos 2005; International Brachypodium Initiative, 2010), they can be used as models for larger plant genomes. Among the model genomes, *Brachypodium* has the most recent divergence time from wheat – 32-39 Million Years Ago (MYA), the smallest genome size and, therefore, it is widely used for gene prediction and construction of synteny-based maps in *Triticeae* (Mayer *et al.* 2011; Wicker *et al.* 2009).

During the past few years wheat genomics has advanced dramatically. The effort to sequence its large genome was divided into smaller steps through construction of physical maps for individual chromosomes. Reducing the complexity of the wheat genome is achieved by isolating single chromosomes or chromosome arms by cytometric flow-sorting (Šafář *et al.* 2010). Nevertheless, there is no reference sequence available for the wheat genome until now. Sequencing of a minimum tiling path of Bacterial Artificial Chromosomes (BAC) clones is under way for several chromosomes (International Wheat Genome Sequencing Consortium www.wheatgenome.org), the most advanced project is the sequencing of chromosome 3B (approximately 1Gb). A physical map for this chromosome was constructed based on a fingerprinted BAC library with 67,968 clones, genetic mapping and deletion-bin mapping (Paux *et al.* 2008). Analysis of collinearity between the physical map of chromosome 3B (cultivar Chinese Spring) and the rice genome has shown that corresponding regions of rice and wheat chromosome 3B are mostly collinear, but there are local rearrangements within this syntenic intervals (Paux *et al.* 2008). Additionally, based on mapping of Expressed Sequence Tags (EST), a positive

gradient of gene density was detected along chromosome 3B from centromere to telomere (Akhunov *et al.* 2003; Munkvold *et al.* 2004). According to the study of Akhunov *et al.* (2003), the recombination rate is higher in distal regions of the chromosome. This finding was strongly supported by Choulet *et al.* (2010) based on annotated megabase-sized fragments of sequence and a recent transcriptional map with 3000-loci of chromosome 3B (Rustenholz *et al.* 2011).

Until now, development of new polymorphic markers was based on very limited resources such as wheat EST data, syntenic information from rice and *Brachypodium* and wheat BAC-sequences from a few loci. New sequencing technologies provide an increase of sequence data from different wheat accessions. SNP assays in combination with new high-throughput genotyping methods are becoming even more attractive for high-resolution genetic mapping and marker-assisted breeding (Agarwal *et al.* 2008). Several recent studies successfully genotyped a few thousand SNPs using genome survey sequences and EST data (Akhunov *et al.* 2009; Barker and Edwards, 2009; Akhunov *et al.* 2010; Allen *et al.* 2011; Chao *et al.* 2009). As modern wheat varieties were bred from a limited number of landraces, the resulting diversity bottleneck lead to modern wheat varieties whose genomes are highly similar (approximately 99.9%) (Ravel *et al.* 2006; Chao *et al.* 2009, Trick *et al.* 2012). For instance, Ravel *et al.* (2006) estimated SNP frequency to be 1 in 335 bp (2.99 SNP/kb), Barker and Edwards (2009) detected 4.29 SNP/kb and Trick *et al.* (2012) found an average density of 1.80 (±1.46) SNP/kb. This low molecular diversity is even more pronounced if only two specific cultivars are compared.

Swiss winter wheat cultivars Arina and Forno were parental lines for a mapping population consisting of Recombinant Inbred Lines (RILs) (Paillard *et al.* 2003). This population is being used for mapping of different valuable traits, including Quantitative Trait Loci (QTLs) for resistance to several fungal diseases. For QTL mapping, genetic linkage maps for all chromosomes of this population were constructed. The genetic map for chromosome 3B had a length of 199.83 centimorgans (cM) and contained 27 markers.

Here we report on the production of a SNP database from flow-sorted chromosome 3B of cultivars Arina and Forno. Seventy SNP-based markers were developed and mapped on the Arina x Forno RIL population. Our analysis suggests that while the gene order is overall well conserved between wheat chromosome 3B and chromosome 2 of *Brachypodium*, there are numerous exceptions in microcollinearity. Additionally, we found that the majority of recombination events occur in distal parts of chromosome 3B. The focused approach described in the present study resulted in a fast development of highly specific polymorphic markers which can be applied in high-resolution mapping and marker-assisted selection.

Results

Purification and sequencing of chromosome 3B

Chromosomes 3B of the two Swiss winter wheat cultivars Arina and Forno were purified by flow cytometry to reduce the sample complexity and in particular to avoid the presence of homoeologous sequences from chromosomes 3A and 3D in the DNA samples. The isolated chromosomes 3B of both cultivars were sequenced by Illumina technology. The number of reads obtained for Arina and Forno were 121,931,740 (approximately 9.6 Gb) and 126,662,154 (approximately 10 Gb), respectively. It has been shown that chromosome 3B has a size of 1Gb (Paux *et al.* 2008). *De novo* assemblies for both cultivars resulted in approximately 300,000 contigs each with an average length of 500 bp. We calculated the average coverage for the contigs containing coding sequences. We produced a graph with the number of contigs plotted against the coverage of those contigs (Figure S1). The peak of the contig number for both cultivars was around 8- to 10-fold coverage (maximum was at eightfold).

Identification of genes based on synteny with *Brachypodium*

To detect putative genes on assembled wheat contigs for both cultivars, we used BLASTN against a database of *B. distachyon* coding sequences. The total number of identified contigs containing coding sequences was 11,563 for Arina and 11,275 for Forno corresponding to approximately 3% of all contigs. This number included different cases. First, if there was exactly one wheat contig matching a specific *Brachypodium* gene, we counted it as a single-copy gene. Second, if two or more different contigs contained homologs of the the same *Brachypodium* gene in a different sequence context, they were considered paralogous copies. Finally, if two or more different contigs contained different parts of the same *Brachypodium* gene with only a slight or no overlap, they were considered to represent parts of the same gene. From this data we defined a “non-redundant” gene set for chromosome 3B. This gene set represents a list of all *Brachypodium* genes that have homologs in the 3B contigs, regardless if several 3B contigs hit the same *Brachypodium* gene. For example, the coding sequence of *Bradi2g49380* was detected on 4 different Arina contigs (indicating a gene family with several paralogs), but in the “non-redundant” set, is was counted only once.

Among all identified gene-containing contigs we detected 4,542 non-redundant genes for Arina and 4,293 for Forno. The number of non-redundant genes included all different *Brachypodium* gene hits associated with sequenced wheat contigs. Although both cultivars had a very similar number of non-redundant genes, we detected 249 non-redundant genes more for cultivar Arina. The distal parts of *Brachypodium* chromosome 2 are syntenic to

wheat chromosome 3B: *Bd2g00200 – Bd2g14080* and *Bd2g40150 – Bd2g62810* (Figure 1a, green bars). About half of the discovered genes had their closest homolog in the syntenic region of *Brachypodium* chromosome 2: 2,333 putative genes (51.4%) for Arina and 2,131 (49.6%) for Forno. Overall, there are 3,674 genes on chromosome 2 of *Brachypodium* in the region syntenic to wheat chromosome 3B. These 3,674 genes have potential homologs on chromosome 3B of wheat. Our analysis detected 63.5% and 58% of these genes in the contig sequences of chromosome 3B from cultivar Arina and Forno, respectively. The number of gene sequences which have their best homologs on the nonsyntenic chromosomes of *Brachypodium* were very similar for both cultivars (Table 1). Additionally, as depicted on the heat-map (Figure 1a), the localization of these sequences does not show any clustering pattern. This observation suggests that contamination from the flow-sorting procedure was low, and that all wheat chromosomes were equally contributing to the contamination.

Table 1. Number of wheat chromosome 3B genes on *Brachypodium* chromosomes identified in sequence contigs of cultivars Arina and Forno

***Brachypodium* chromosome**	**Arina**	**Forno**
1	764	743
2a	2,333	2.131
2b	231	218
3	567	560
4	391	391
5	249	244

2a - genes from syntenic part of chromosome 2 of *Brachypodium*
2b - genes from nonsyntenic region of chromosome 2

Differences in gene content between Arina and Forno

The combined number of non-redundant genes detected for chromosome 3B of both cultivars was 5,224. However, 18% (931) of the genes were present only in Arina and 13% (682 genes) were present only in Forno (Figure 1b). Therefore, approximately 30% of all putative genes detected in the cultivars were different. A very similar result was found for the syntenic genes: 2,007 of them were common to both cultivars and 326 putative genes were present only in cultivar Arina while 124 were present exclusively in cultivar Forno (Figure 1c).

To check whether we can use these data to identify presence/absence polymorphisms in gene content of Arina and Forno, we designed 67 primer pairs from genes specifically found only in Arina or Forno chromosome 3B. We used DNA of wheat cultivars Arina, Forno, Chinese Spring and the Chinese Spring nulli-tetrasomic line (N3BT3D) for amplification of fragments. The last two lines were included as a control for chromosome 3B specificity of the amplicons. Amplification was successful in 45 cases but only five fragments were specific to the 3B chromosome (no amplification from the Chinese Spring nulli-tetrasomic line was obtained, whereas in Chinese Spring a PCR product was present). The large number of amplified fragments which are not specific for chromosome 3B indicates that these primers also amplified homoeologous sequences. Three of the five chromosome 3B-specific sequences (InDel_g01210, 512-bp-long fragment amplified from Forno, InDel_09780, 143 bp amplified from Arina and InDel_56500, 818 bp fragment amplified from Forno) were polymorphic between Arina and Forno and were mapped on the genetic map (see below). Therefore, out of 5 fragments specific to chromosome 3B, 3 were confirmed to be useful as presence/absence polymorphisms.

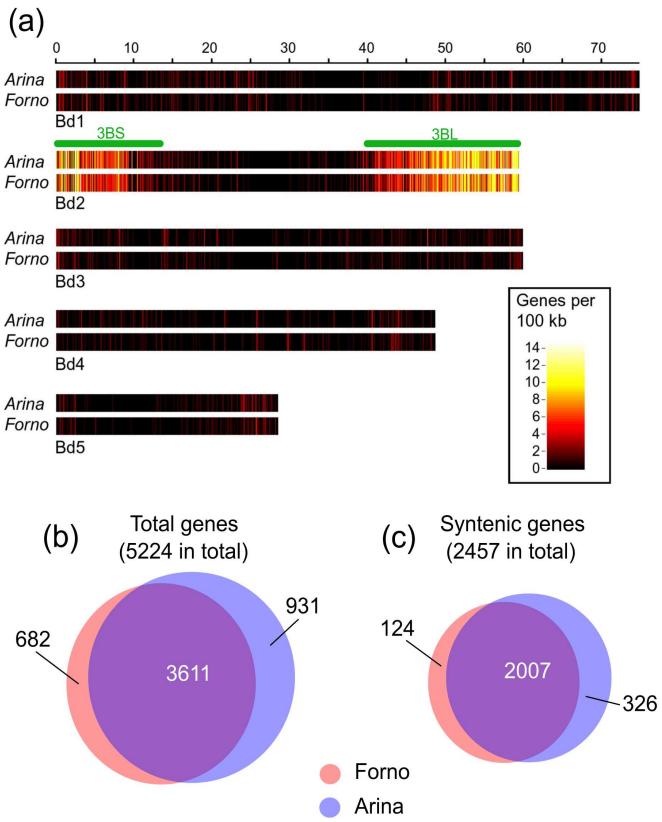

Figure 1. Comparison of identified genes on chromosome 3B of bread wheat cultivars Arina and Forno

(a) Heatmap depicting the distribution of homologs of wheat chromosome 3B genes in the *Brachypodium distachyon* genome. Bd1 through Bd5 correspond to *B. distachyon* chromosomes 1 through 5. The scale at the top is in Mbp. Gene-containing sequence contigs from Arina (top) and Forno (bottom) were mapped independently to the *Brachypodium* genome. The regions syntenic to wheat chromosome 3B have the highest density of homologs. Signals in nonsyntenic regions suggest a presence of true nonsyntenic genes if they occur in both Arina and Forno. Green bars correspond to the syntenic regions of wheat chromosome 3B. **(b)** Venn diagram with all detected non-redundant coding sequences in each cultivar. **(c)** Diagram depicts the number of shared and unique genes in the two cultivars with a *Brachypodium* gene homolog from the syntenic region (distal parts of chromosome 2 of *Brachypodium)*.

Identification of SNPs in gene-containing contigs

We chose the Arina *de novo* sequence assembly as a reference on to which Forno Illumina reads were mapped for the identification of polymorphisms. Nucleotide variations between two cultivars were identified using the CLC Assembly Cell (Figure 2). In total, 1,835,214 SNPs were detected. To reduce the problem of unspecific comparisons mostly in repetitive DNA regions, only contigs containing coding sequences of genes were chosen for further analysis. Gene-containing contigs were selected for further analysis if they had a unique gene hit or if a contig did not have any overlaps with other contigs which had the same gene hit. This selection step was made to exclude paralogous gene copies for marker development.

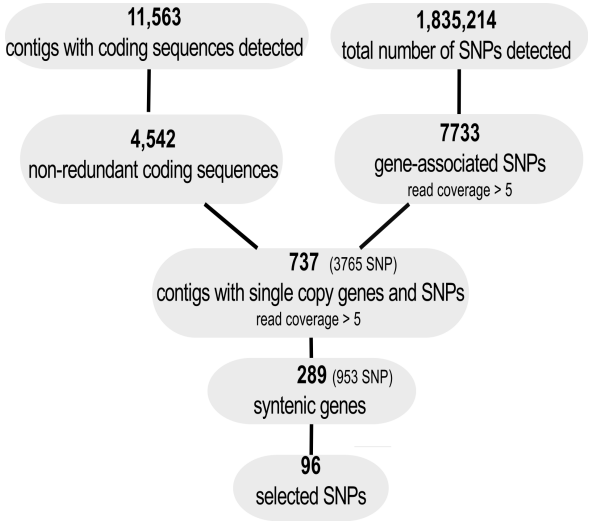

Figure 2. Scheme of SNP selection steps

Coding sequences (CDS) on the contigs of Arina assembly were detected using a database of *Brachypodium distachyon* coding sequences. The number of non-redundant coding sequences included all different *Brachypodium* gene hits associated with sequenced wheat contigs.

A further selection among the contigs containing suitable coding sequences and SNPs was made to reduce the number of SNPs caused by sequencing errors. First, we applied a sequence coverage threshold for each candidate SNP. Nucleotide polymorphisms in regions with a coverage below five reads were not considered for SNP marker development (Figure 2). Second, we mapped all Arina reads to the reference Arina assembly in order to control base calling and to eliminate candidate SNP positions with low sequence coverage. All SNPs with a coverage below five in Arina or ambiguous base calling were excluded. In total, 737 contigs with high-quality SNPs were selected. Among them, 448 contigs had gene hits from nonsyntenic regions of *Brachypodium* and 289 were from the region which corresponded to chromosome 3B. From them, 131 belonged to coding sequences from the short chromosome arm and 158 contigs had gene hits corresponding to the long arm of 3B.

Development of a cultivar-specific genetic map of chromosome 3B based on 70 SNP markers

For SNP mapping, we used only contigs representing wheat genes which have their homologs in the 3B syntenic regions of *Brachypodium*. We chose genes expected to be regularly spread across the whole chromosome based on syntenic location in *Brachypodium*. In total, 96 SNPs were selected for the genetic mapping on 178 RILs derived from a cross between cultivar Arina and cultivar Forno (Paillard *et al.* 2003).

A KASPar assay was chosen for genotyping of the 178 RIL plants. Before designing primers for the KASPar assay, we masked the coding parts of the selected contigs. Only SNPs located upstream or downstream of the gene's CDS were selected. This strategy aimed at reducing the possibility that homoeologous or paralogous copies of targeted genes were amplified. As a result, 87 SNP containing fragments were amplified, among them 64 were polymorphic and suitable for genetic mapping. In addition to these 64 markers, 6 polymorphic SNPs from 8 tested were mapped on the RIL population in a pilot experiment to test the feasibility of the assay, resulting in a total of 70 markers derived from the SNP assay . Therefore, in total 104 SNP markers were tested, 95 of them amplified and 70 markers were mapped (Table S1).

The initial genetic map of chromosome 3B of Arina x Forno RIL population consisted of 27 markers and had a size of 199.83 cM. After the integration of 70 additional SNP and 3 InDel markers the length of the map was 202.2 cM (Figure 3). This indicated that the integration of the added markers was of good quality. The names of new SNP-based markers were chosen according to the *Brachypodium* gene homologs which were present on the wheat contigs that contained the SNPs (e.g. *Brachypodium* gene *Bd2g47720* – wheat marker g47720).

To test whether the SNP markers specifically developed from Arina and Forno were applicable in a larger set of genetic material, we tested a subset of 48 SNP markers on an additional 44 wheat varieties with different origins: European spring and winter wheat germplasm, US germplasm, Mexican varieties and Chinese landraces (Table S2). The test revealed a presence of both Arina and Forno alleles with variable frequencies from 0.02 to 1 for each allele (Figure S2). Among the 48 SNPs only one SNP represented a private allele of cv. Forno, whereas all other SNPs were present in a least one of the tested lines. Thus, the developed SNPs are widely polymorphic in germplasm and add to the available SNP markers in wheat.

SNP mapping reveals overall good colinearity with *Brachypodium* with several exceptions

We expected to find a good conservation between gene order on chromosome 2 of *Brachypodium* and wheat chromosome 3B. Indeed, the marker order on the newly constructed wheat map was mostly consistent with the gene order found on chromosome 2 of *Brachypodium*. Overall, only eight of the 73 (70 SNP and 3 InDel markers) analyzed markers were in nonsyntenic positions. As depicted in Figure 3, we observed co-segregation of a few groups of markers, especially in the centromeric region. With a few exceptions, these groups were formed by markers with their *Brachypodium* gene homologs in syntenic positions. For example, the largest group of 12 co-segregating markers included only one nonsyntenic marker, g62500 (Figure 3, cluster B in cyan). The 11 syntenic genes in this cluster (*Bd2g07680 – Bd2g09470*) span a region of 1.73 Mb in *Brachypodium*. In wheat, the corresponding gene-based SNP markers co-segregated on the genetic map in a region close to the centromere.

We detected four inversions on the genetic map of chromosome 3B by comparison with the syntenic order of the genes on *Brachypodium* chromosome 2 (Figure 3, clusters A, C, D, E). The sizes of the identified inversions ranged from 0.6 cM and 1.3 cM in proximal regions to 7.6 cM in the telomeric part. One of these inversions (cluster C) consisted of 9 genes (*Bd2g46680 – Bd2g47750*) located in a 1 Mb interval on *Brachypodium* chromosome 2 (Figure 3, cluster C in cyan). In wheat, this inversion comprised a genetic interval of 1.2 cM. Such information on breaks in synteny and local, small-scale inversions will be essential for high-resolution mapping projects.

SNPs are unevenly distributed along chromosome 3B

To analyze the distribution of gene-associated SNP markers on the genetic map of chromosome 3B, we examined the correlation between the distribution of the 289 high-

quality, syntenic, gene-associated SNPs and the distribution of all syntenic *Brachypodium* gene homologs detected on wheat contigs (2,333 Arina genes) on chromosome 2 of *Brachypodium*. We mapped *in silico* all identified syntenic genes from wheat contigs to their homologous positions on the chromosome 2 of *Brachypodium*. The correlation was tested by calculating the ratio between the number of selected genes with SNPs and the total number of identified syntenic genes. The ratio was computed using a sliding window approach: we took intervals of 1.2 Mb with incremental steps of 100 kb (Table S3). The ratios, obtained for each interval, are shown in Figure 3 (blue graph). This analysis revealed that the type of SNPs studied here were distributed unevenly: the number of genes with SNPs, normalized to the total amount of genes, had a few peaks and was generally higher in the region corresponding to the short arm of chromosome 3B. This suggests that the frequency of gene-associated SNPs in cultivars Forno and Arina is higher in 3BS than in 3BL.

Low recombination in the centromeric region of chromosome 3B

We also wanted to compare the genetic distances on the map of wheat chromosome 3B with the physical size and number of syntenic genes in the corresponding chromosomal regions in *Brachypodium*. We focused our analysis on a region of 18 cM which included the centromeric region of the map with a large gene cluster (Figure 3, purple block). There were a total of 47 mapped gene-associated SNPs and 1255 syntenic *Brachypodium* genes detected in this region based on the sequences from wheat cultivar Arina. The wheat region of 18 cM around the centromere corresponded to a segment on *Brachypodium* chromosome 2 which spanned a total size of 19.34 Mb, with 9.23 Mb on the syntenic region of the short arm and 10.11 Mb on the syntenic region of the long arm. Therefore, in this centromeric region 53.79% of all syntenic genes (1255 of 2333) were found in a region corresponding to 8.9% (18 cM) of the total genetic distance in wheat and 62.89 % of the total physical size 30.75 Mb of the syntenic region in *Brachypodium*.

Brachypodium genes homologous to the gene-associated SNP markers in the distal parts of the genetic map were assigned to the corresponding yellow blocks (Figure 3). Physical intervals corresponding to the distal parts of the genetic map had sizes of 2.63 Mb for the short arm of chromosome 3B and 8.78 Mb for the long arm of chromosome 3B (together corresponding to 37.11% of the 30.75 Mb physical region syntenic in *Brachypodium*). The two distal parts of the genetic map comprised 91.1% of the total map size in wheat (184.2 cM) and contained 46.21% (1078) of detected genes. These observations demonstrate that in Arina x Forno population there is a strong suppression of recombination in the centromeric region of wheat chromosome 3B, but also that there are many genes in this region.

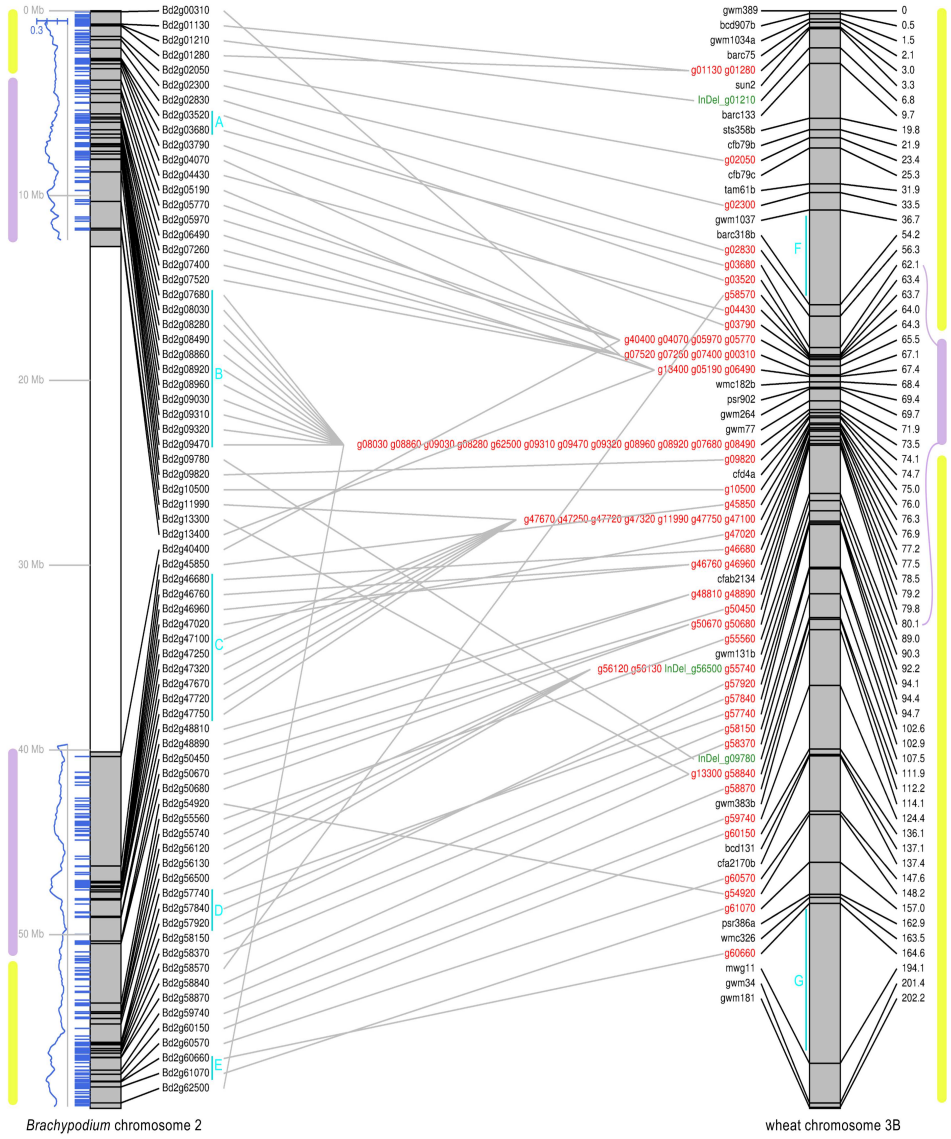

Figure 3. Comparison of the genetic map of wheat chromosome 3B of RIL Arina x Forno population and the physical map of *Brachypodium* chromosome 2

In gray: syntenic regions of *Brachypodium* chromosome 2 and wheat chromosome 3B. **In red:** newly mapped gene-based SNP markers. Marker names correspond to the gene homologs in *Brachypodium*. Genetic distances are given in centimorgans (cM). **Blue bars:** physical positions of *Brachypodium* gene homologs corresponding to the selected gene-associated contigs with SNPs of wheat chromosome 3B; gene homologs that were mapped as SNP markers are represented both as a blue bar and black bar on the *Brachypodium* chromosome. **Blue graph** depicts a ratio between the number of selected wheat gene-containing contigs with SNPs and the total number of identified non-redundant *Brachypodium* genes in wheat sequenced contigs. The ratio was calculated using a sliding window of 1.2 Mb and incremental step 100 kb. **In cyan: A, C, D, E** - regions of detected inversions between wheat and *Brachypodium,* **B** is a cluster of mapped *Brachypodium* genes, regions F and G – gaps on genetic map. **In purple and yellow:** centromeric region of genetic map (purple) and corresponding physical region on *Brachypodium* chromosome 2; distal parts of genetic map with corresponding physical *Brachypodium* regions (yellow). **In green**: marker derived from gene presence-absence polymorphism.

Gaps on the genetic map

The Arina/Forno SNP map contained two relatively large regions not covered by genetic markers (Figure 3, region F and region G). The region G (*g60660 – mwg11*) has a size of 29.5 cM and the last gene-associated SNP mapped to this area (*g60660*) is 36.7 cM away from the telomeric end of the genetic map. The number of detected syntenic *Brachypodium* gene homologs in the Arina dataset was 115 for the region from *Bd2g61070* – end of the *Brachypodium* chromosome 2. Therefore, further mapping could probably place some of these 115 genes in this gap. Additionally, Saintenac *et al.* (2009) showed that recombination frequency in this region is very high. Furthermore, other genetic maps indicated a similar absence of markers: on the ITMI composite map, marker wmc326 (on our map it is placed 1.1 cM proximal from the described interval) is located at position 114 cM while the total size of the map is 148 cM (Somers *et al.* 2004). On the map of double haploid lines Avalon x Cadenza, wmc326 was mapped to position 138.8 cM (total size of the map was 183.0 cM). In our case, wmc326 is located at 163.5 cM position (total size of the map is 202.2 cM). Gaps are representing the regions where genetic distance is large (a high level of recombination), but the number of markers is small. Based on the syntenic comparison, there is no evidence for large physical regions in the corresponding regions of wheat. Therefore, the described gap *g60660 – mwg11* can most likely be explained by a high recombination level in this region.

The region F (*gwm1037 – barc318b*) is 17.5 cM long and does not contain any genetic markers. We checked the number of *Brachypodium* gene homologs which were detected in Arina in the interval *Bd2g02300 – Bd2g02830 (*g02300 is 3.2 cM distal from *gwm1037* and *g02830* is 2.1 cM proximal from barc318b). There are 31 *Brachypodium* genes detected. Hence, this gap can possibly also be explained by the presence of a recombination hot-spot or by nonsyntenic genes intercalated in this region.

Discussion

Gene-associated SNP markers based on whole chromosome sequencing are highly specific

Mapping in hexaploid wheat is complicated by the difficulty to derive specific markers because of the highly repetitive genome and the presence of both paralogous and. homoeologous copies of a particular gene in the three subgenomes. Therefore, an essential part of any mapping strategy is to increase marker specificity. In the present study we used sequences from isolated chromosome 3B. Flow cytometric analysis of DAPI-stained mitotic

chromosomes of wheat cv. Chinese Spring results in a histogram of relative fluorescence intensity (flow karyotype) comprising three composite peaks of groups of chromosomes and a peak corresponding to chromosome 3B (Vrána *et al.* 2000). Therefore, flow sorting of chromosome 3B is easier than the isolation of other wheat chromosomes. These can be sorted as chromosome arms from cytogenetic stocks in which the arms are stably maintained as telocentric chromosomes (Doležel *et al.* 2007). Alternatively, fractions highly enriched for target chromosomes and free of homoelogous chromosomes can be obtained by applying narrow sort windows (J. Doležel, unpublished data). Any of these strategies reduces DNA sample complexity, which facilitates efficient and targeted development of genetic markers (Wenzl *et al.* 2010).

The main strategy of our study was to use gene-associated SNPs to improve marker specificity for chromosome 3B and to get an even distribution of markers based on gene synteny with *Brachypodium*. To gain specificity for designed SNP markers both between homoeologous chromosomes and within chromosome 3B, we used a set of criteria for selection of SNP. All repetitive sequences were excluded from the analysis, because amplification from repetitive regions could lead to unspecific amplification from similar repetitive sequences. Furthermore, a limitation of the sequence read coverage threshold for each SNP position was established. We selected only contigs with non-redundant gene hits and analyzed SNPs located outside of the coding region. This setup resulted in a low number of nonspecific markers and a final success rate of approximately 74% (70 out of 95) for SNP marker mapping. We used very stringent parameters for SNP selection (e.g. only syntenic genes, only one contig per given gene, read coverage greater than 5, the SNP had to be present in all reads). It is therefore possible that more flexible parameters might give a higher yield of SNPs for mapping. Nevertheless, we consider the number of identified SNPs and the yield in marker development a success, because low polymorphism is a general problem in wheat genetics.

Several prior studies showed that wheat has a low level of varietal SNP polymorphisms useful for marker development (Barker and Edwards 2009). A recent study by Allen *et al.* (2011) revealed that among 1,659 SNPs selected using a cDNA library of five wheat varieties, 67% SNPs were polymorphic between different varieties, but only 10% of polymorphic SNPs showed genome specificity. The approach for optimization of genome specificity, suggested in the study by Allen *et al.* (2011), included re-design of KASPar probes based on alignments of three genome-specific sequences for corresponding regions from the Chinese Spring 5-fold sequence database (www.cerealsdb.uk.net). They report that their approach indeed increased genome-specificity of SNPs, but the method is time consuming and laborious. Trick *et al.* (2012) mapped 39 SNPs on tetraploid wheat using RNAseq and wheat unigenes. Discrimination between the copies on the two different genomes was done by alignment of sequences to unigenes and the design of genome-specific primers for KASPar which led to a success rate of 74% in this study.

The strategy used in our study resulted in high percentage of polymorphic SNP markers. The vast majority of produced SNP markers (66 out of 70) were genome specific. This level of efficiency is high given the problem of three closely related homoeologous genomes in wheat and was obtained without any additional steps such as genomic alignments and re-designing of primers for the KASPar assay. Furthermore, preselection of potential SNPs based on syntenic gene homologs allowed us to obtain a more predictable distribution of SNP markers on the genetic map. Preselection could be particularly advantageous for high-resolution mapping projects.

Wheat and *Brachypodium* show a high degree of synteny but also many exceptions

We used a synteny-based approach to identify putative genes on sequenced wheat contigs. Because *B. distachyon* is the closest wheat relative with a small and entirely sequenced genome, identification of wheat genes and their order based on the synteny with *Brachypodium* have been used extensively (Bossolini *et al.* 2007; Quraishi *et al.* 2011). Our analysis revealed that about half of the identified genes (2,333 from 4,542 in Arina and 2,131 from 4,293 in Forno) had their homologs in distal parts of chromosome 2 of *Brachypodium*, the 3B syntenic regions. Additionally, the absence of large gene clusters with corresponding gene homologs on other *Brachypodium* chromosomes indicated an absence of large chromosome translocations. However, there is still a possibility of large translocations in gene-poor regions which cannot be excluded with our approach. Genetic mapping of gene-associated SNPs also revealed several inversions between gene order on *Brachypodium* chromosome 2 and the genetic map of chromosome 3B (Figure 3, cyan bars).

Gene content of chromosome 3B

Besides the syntenic genes, a large number of nonsyntenic genes were identified. This observation is consistent with previous findings about gene content of wheat chromosome 3B. It was suggested that nonsyntenic genes in wheat are interspersed within a very conserved syntenic gene set (Choulet *et al.* 2010). Indeed, we found that roughly 50% of the detected genes mapped to nonsyntenic regions in *Brachypodium* in both Arina and Forno assemblies. Several recent studies provided estimates of the total gene number for chromosome 3B between 6,000 and 8,400 genes (Paux *et al.* 2006; Choulet *et al.* 2010). Our observations are slightly different: we detected roughly 11,000 gene-containing contigs in each cultivar. However, since we did not conduct a detailed gene annotation, this number possibly includes many pseudogenes, gene fragments and misassembled parts

of the same genes. Nevertheless, we were able to identify a minimal gene set for chromosome 3B consisting of approximately 3,600 non-redundant genes (Figure 1b). This minimal set of genes was detected independently in both Arina and Forno datasets, suggesting a high accuracy of gene prediction for this set.

In Arina 14% of genes with synteny to *Brachypodium* chromosome 2 were cultivar-specific, in Forno 5.8%. These numbers could be explained by the presence of real gene deletions in Arina and Forno genomes or by insufficient sequence coverage of the samples. We could confirm some presence/absence polymorphisms by mapping of InDel markers, but most of the other amplification products were not chromosome 3B specific, probably due to homoeologus sequences in the A and D genomes. The calculated coverage for our dataset is 8 fold and we expect 99.97% of the genes to be sampled (Lander and Waterman 1988). However, a purely random coverage can not be expected: A certain bias in the sequence coverage has been shown to be caused by multiple displacement amplification of flow-sorted chromosomes and the sequencing itself (Pinard *et al.* 2006; Wicker *et al.* 2011). Additionally, the high number of cultivar-specific genes could be caused by our gene annotation procedure, which was based on best homology to *Brachypodium* genes. For example, it is possible that sequence contigs from Arina and Forno cover different parts of a given gene. Homology search in *Brachypodium* could then result in two different genes as best hit, wrongly indicating that both genes are cultivar specific.

The level of recombination varies strongly along the chromosome

The majority of mapped markers were placed in the proximal centromeric region of chromosome 3B. In addition, numerous markers from the centromeric region co-segregated genetically. The proximal centromeric region of the genetic map spanned 18 cM (less than 10% of the genetic map) and included 1,255 genes in the syntenic regions of *Brachypodium* (53.8% of the genes from the syntenic regions of *Brachypodium* chromosome 2). The rest of chromosome 3B (distal regions) had a genetic size of 62.1 cM and 122.1 cM, and included 210 (9.0%) and 868 (37.2%) of the genes in the syntenic regions of chromosome 2, respectively. This observation demonstrates that recombination was highly suppressed in the centromeric region. This finding is consistent with previous studies: Saintenac *et al.* (2009) analyzed crossover frequency using data from physical and genetic maps of wheat chromosome 3B. They detected a crossover frequency of 90% in distal subtelomeric parts and a very low recombination in the proximal centromeric region of chromosome 3B. The comparison of the chromosome 3B genetic map and gene content as estimated by comparison with syntenic chromosome 2 of *Brachypodium* very well reflected the detected low recombination rate around the centromeric region (Figure 3, purple and yellow blocks). This fact suggests that mapping in this region is especially challenging.

Differences in gene content between two cultivars: contamination or tool for marker development?

The absence of large clusters of *Brachypodium* gene homologs elsewhere than in the syntenic regions of *Brachypodium* chromosome 2 indicated that the contamination of the chromosome 3B fraction with other chromosomes during flow cytometric sorting was low. Furthermore, a maximal degree of contamination could be estimated based on a total number of genes exclusively present in Arina or in Forno (Figure 1b). The detected number of cultivar-specific genes in Arina and Forno was 931 and 682, respectively. If we assume that identification of those genes was caused entirely by contamination during the flow-sorting, the maximal estimation would be 20% of contaminated fraction for Arina and 16% for Forno. This is more than estimated by Fluorescent In Situ Hybridization (FISH) on fractions of sorted chromosomes (3.2% and 8.9% for Arina and Forno, respectively). However, if the detected variation in gene content between cultivars Arina and Forno was based not solely on contamination, it has at least two potential applications. First, genes which are absent in one cultivar and present in another could be used as dominant presence/absence markers for genetic mapping. This concept is supported by the fact that we could map 3 of 5 chromosome 3B specific gene-based InDel markers (Figure 3, InDels in green). However, the development of such markers is inefficient as the primers developed from presence/absence polymorphism are mostly not chromosome-specific and also amplify homoeologous sequences. A second application of InDel variation could be in candidate gene selection in fine mapping studies.

Whole chromosome sequence-based mapping in wheat

The strategy based on sequencing flow-sorted chromosomes can be used for targeted genotyping of different wheat varieties or populations both in high and low throughput manner. We selected SNPs that were gene-associated, but not positioned in the coding sequences themselves. This factor potentially provides a higher rate of polymorphisms between elite wheat varieties in comparison with SNPs in coding regions. The developed SNP markers are genome-specific and therefore can be used not only for screening of homozygous plants but are also applicable for mapping in heterozygous populations. The analysis of a subset of 48 SNP-based markers, specifically developed for cultivars Forno and Arina,in a broader set of 44 genotypes allowed us to determine the frequency of polymorphisms in such germplasm. The results suggest that developed SNP can be widely used in the genepool. In conclusion, SNP markers are becoming increasingly important for wheat genetic mapping and marker-assisted selection. The strategy to develop SNP markers applied in the present study is a fast and efficient approach to genetic mapping. A combination of sequencing of individual chromosomes or chromosome groups and

selection of SNPs associated with genes, but located outside of the coding sequences produced highly specific results.

Experimental procedures

Plant material

Two varieties of winter wheat (*Triticum aestivum* L., 2n=6x=42) 'Arina' and 'Forno' were chosen for SNP analysis. Genetic mapping was done on a population of 178 RILs, derived from a cross of cultivar 'Arina' and cultivar 'Forno' (Paillard *et al.* 2003). DNA of cultivars 'Arina', 'Forno', 'Chinese Spring' and nulli-tetrasomic chromosome 3B line N3BT3D of 'Chinese Spring' were used for InDel PCR analysis. DNA extraction from leaf material was done as described in Stein *et al.* (2001).

Flow-sorting of chromosome 3B

Purification of chromosome 3B was done by flow sorting. Liquid suspensions of mitotic chromosomes were prepared from synchronized root tips of both wheat cultivars according to Vrána *et al.* (2000). Chromosome samples were stained by DAPI and chromosome 3B was isolated by flow cytometric sorting as described by Kubaláková *et al.* (2002). The chromosomes were sorted in several batches of 25,000 chromosomes with the purity of 96.8% and 91.1% for Arina and Forno, respectively, as determined by FISH with probes for GAA and *Afa* repeat. DNA of isolated chromosomes was amplified *via* multiple displacement amplification (MDA) using Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare). Two and four independent amplifications were made for Arina and Forno respectively. In total, 8.61 μg and 10.92 μg DNA were obtained for Arina and Forno, respectively.

Sequencing and assembly of chromosomes 3B of cultivars Arina and Forno

Sequencing of chromosomes was performed by GATC Biotech, Konstanz, Germany. For each cultivar, 5 μg DNA was sequenced by Illumina technology with independent libraries of 200-250 bp sequence fragments and paired ends. After quality trimming, 100 bp long Illumina reads of Arina and Forno were separately assembled *de novo* using clc_novo_assemble command with default parameters of CLC Assembly Cell 3.22 software (CLC bio, Aarhus, Denmark).

Calculation of expected gene coverage

To calculate the expected gene coverage we used the following formula based on the Lander-Waterman approach:

$P=1-e^{-c}$, where P is a probability that any base in contigs with genes is covered and c – average coverage of the contigs with coding sequences. The average coverage was calculated as described above (Figure S1).

Identification and selection of SNPs

Forno reads were mapped to the Arina *de novo* assembly using function "clc_ref_assemble_long" of CLC Assembly Cell. Potential SNPs were identified using function "find_variations" of CLC Assembly Cell. Read coverage for each base pair position in targeted contigs was extracted using "assembly_info" CLC Assembly Cell. BLASTN against *Brachypodium distachyon* coding sequences database was used to detect putative genes on wheat contigs. Perl scripts were written to mine and extract the data for SNP selection and are available upon request.

RIL genotyping and construction of the genetic map

One hundred seventy eight RILs Arina x Forno were chosen for testing of 104 identified SNPs. Genotyping was performed by KBioscience (Herts, UK) using fluorescence-based competitive allele specific PCR (KASPar) assay. The genetic map was constructed using MultiPoint software (Mester *et al.* 2003). Distances were calculated using Kosambi function (Kosambi 1944).

Acknowledgements

We thank Bea Senger for technical assistance with the plant material, Gerhard Herren for the support with the PCR analysis and Roi Ben David for the advice on construction of the genetic map. The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the grant agreement FP7-212019 (Triticeae Genome) and was supported in the framework of the European Cooperation in Science and Technology FA0604 (Tritigen). This work was also supported by grant a from the Swiss National Science Foundation 31003A_127061.

Supplementary materials

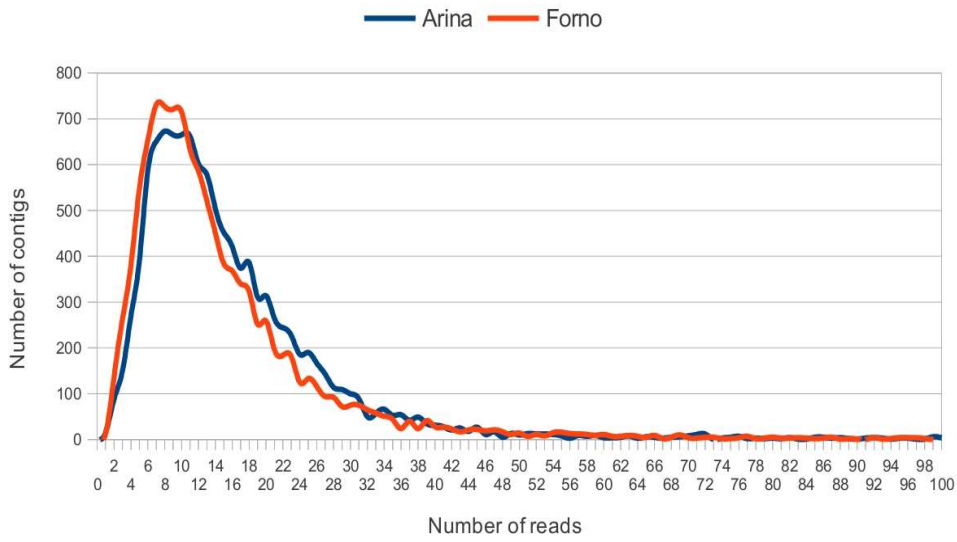

Figure S1. Sequence coverage of contigs with coding sequences

The average sequence coverage of the contigs with coding sequences has a bell-shaped distribution with a long tail. The highest number of contigs has a coverage of 8 reads for both Arina and Forno.

Table S1. Primer sequences for developed SNP and InDel markers

	Primer Allele Arina	Primer Allele Forno	Primer Common	Allele Arina	Allele Forno
g00310	TCTTGGAAGTGTAGCGCCATGTA	CTTGGAAGTGTAGCGCCATGTG	GCAAGCCTCTCAACATTGCTACCAT	A	G
g01130	CATAGCTYCATCTCATTTCCTCCTTA	ATAGCTYCATCTCATTTCCTCCTTG	TAGAACCAAAGCKTCAAACATTTCTGTGAA	A	G
g01280	GTAGGTCTCTAAAGGAACTG	GCTGTAGGTCTCTAAAGGAACTC	TTGGAGTAAACCTCTACCTCTKAACTATAT	C	G
g02050	ACTACGAGAAGCAGGAGAGTCTT	CTACGAGAAGCAGGAGAGTCTC	TGCTCTCCATCTCGCTGAAATCCTT	T	C
g02300	AGGGTGTCTCTTTCAGAAAACASTAAT	GGGTGTCTCTTTCAGAAAACASTAAC	GACACAAGTGTGCACCAGAGAAGAT	T	C
g02830	TCCAATGCGCTAATCCCGTGC	GTTCCAATGCGCTAATCCCGTGT	TTAGATAATCGAGAAGTGGTCGTGCTAAA	C	T
g03520	CACCCAAACAGCCACATTTTGCG	CACCCAAACAGCCACATTTTGCC	CAGGAAGTGCCCTTCTTGACATGTA	G	C
g03680	TCCTTCTAGCTCGTAAATGATTTACG	GTTCCTTCTAGCTCGTAAATGATTTACA	CACACCGACTATCATTTGTATTGCAGAAA	C	T
g03790	GGACTTGTTTTGGGCGACAATTTGT	GACTTGTTTTGGGCGACAATTTGC	GGTTTTGGACGAAATTCGTCCGGTT	T	C
g04070	GCCAAGAAGTACAGGATCAGGC	CGCCAAGAAGTACAGGATCAGGT	CGTGGCGCAACAAGACTAACCTTA	G	A
g04430	AGGCTGGGCATGGCACATACT	GGCTGGGCATGGCACATACG	ATGAGTTCATTCTTGACTTATGCCTGCTA	T	G
g05190	ATAATACTACTATACCAAAACTGATGCACA	ATACTACTATACCAAAACTGATGCACG	CATGCCTTTGATTTAAACAACCTGCAGTT	A	G
g05770	CAGAACGGGTTCAGGGTGCTC	CAGAACGGGTTCAGGGTGCTT	GTACTTCTTGATGGGATAYGGGATGTA	G	A
g05970	GGCAACAGAGTAGTCAATGATAGTTAATT	GCAACAGAGTAGTCAATGATAGTTAATG	CACGGCTGAAAAGGTTCTCATGCTT	T	G
g06490	CATCATCGAAGAATCAGCTCTCC	CCATCATCGAAGAATCAGCTCTCT	TGGATCGGTCTCTCCTCAGGATTTA	C	T
g07260	CAATCTTTGAGCACGGCGTCA	CAATCTTTGAGCACGGCGTCG	CGTCTGATCCATGCCAACCTTGAAT	A	G
g07400	CAGAAGGATCCTGCTAGGGTATTAAA	AGAAGGATCCTGCTAGGGTATTAAG	CAAGAAGTATGCGCCGTGTGTTCAT	T	C
g07520	TCTAACTGGGAGTTCCAGCTC	CTTCTAACTGGGAGTTCCAGCTT	CCGCCCGAAAATGAGAATTGCCTT	G	A
g07680	TGTAGCAATACTTCCAAACTGGC	CTTGTAGCAATACTTCCAAACTGGG	GCTGAGGCTAAGGGAGAAGTACTT	C	G
g08030	CCGTCACTTGTAGATCAGGAAAAG	GTCCGTCACTTGTAGATCAGGAAAAA	CTGCGCTGTTTGAGAATACAGAACAAAA	G	A
g08280	GTGTAGGACTTGTGCTTTATCTTAAAG	TGTGTAGGACTTGTGCTTTATCTTAAAA	TGTATTTAATGGGGCTTAATCTTGTGGTAA	G	A
g08490	AGCATCCATACATCGGTTACCATGT	CATCCATACATCGGTTACCATGC	CATAGTTCACCTGTATTCCTATTGAGGTT	T	C
g08860	GCACAGGATTCGTTAGAGGG	GCTGCACAGGATTCGTTAGAGGA	TTGCTGGCAGCACTCARGAGCTTA	G	A
g08920	GGGAACATTGGGTCAACACCTG	GGGAACATTGGGTCAACACCTC	CACGGACTTGGTAACCTCCATCTTA	C	G
g08960	TTTGGAATATTTGTAAGCATTATGACTTC	CCTTTTGGAATATTTGTAAGCATTATGACTTA	CATAGGCACCCCAACTGGGCAA	C	A
g09030	TTGGAAAATCAAGCATGCTCGTCATT	GGAAAATCAAGCATGCTCGTCATC	AACAGATCTTAAAGCGGCAAACACATCAT	A	G
g09310	AGGCACAACCAATCGTATATAACATG	AGGCACAACCAATCGTATATAACATC	GACTTGATTTGCCATGCAGGGTTTGTT	G	C
g09320	CATGCATGTGTCTATATACTCAATTTCTTT	ATGCATGTGTCTATATACTCAATTTCTTC	AAAGTCTGCGTGGTAYAGTGCGGA	T	C
g09470	GAATATTTCCTCCATGGTTGTCTGTA	GAATATTTCCTCCATGGTTGTCTGTG	GCACTTGTGGTTCCAAAGATAAACAGAAA	A	G
g09820	GTGTTATCAGTTATCTGGCCTAAAGTT	GTTATCAGTTATCTGGCCTAAAGTC	GTGTCACARCTTACAAAAACTTGTTGGATA	A	G
g10500	CGTGGATCTTGTCTTCATTTCAAGCA	GTGGATCTTGTCTTCATTTCAAGCG	CAATCCATGCATAAAATCTTCAGTCATCAT	T	C
g11990	AATCTGACAACCATACCCATCCC	CAATCTGACAACCATACCCATCCT	GCTCTCGTATAATCTACTTAGGCTGTTAT	C	T
g13300	CAGGTTTCTTTGTAAGCAGGCAG	CAGGTTTCTTTGTAAGCAGGCAC	ACGCAGGAATGAGCAGTTCAGCTTA	G	C
g13400	CATTGGATGCCGCAATTACATCACTA	CATTGGATGCCGCAATTACATCACTT	CCCCTAATAGGATTACCTGATGCGAT	A	T
g40400	ACTTCTTTCACACTTGACAAATTGTTATAAAT	CTTCTTTCACACTTGACAAATTGTTATAAAC	GCACTGATCAGATCAAAGTACATACTAATA	T	C
g45850	AAATTTTCATCCTACGATCATGATCCC	CAAAATTTTCATCCTACGATCATGATCCT	CCAACTCGTTGTTGGTAACAATCTGATTT	C	T
g46680	CCAGATCAAACAATCTACTCCCG	CCCAGATCAAACAATCTACTCCCA	GTCGCCAAGTGGGATCAAAGGTTAA	C	T
g46760	GCCATGAGAAAAAGCGGCACTG	AGCCATGAGAAAAAGCGGCACTT	GGCTTGGCCCCCCTATACCTAA	C	A
g46960	GATGCAGAATCTGTTCCTGGTACA	GATGCAGAATCTGTTCCTGGTACT	TCAGTAATTTGAGAGCAGTGACTTCCATT	A	T
g47020	CTTCGCCGCGGCCAACGA	CTTCGCCGCGGCCAACGC	GTTGCCGCAGGTGGATGGGTA	A	C
g47100	AAAGCAAGGTATTTGGTTAGTTTTCTCTTT	GCAAGGTATTTGGTTAGTTTTCTCTTC	GCTTTGGAACAAACAAAAGGACGATTCAT	T	C
g47250	GATCTTGGTTGGGTACCAGATTCAA	ATCTTGGTTGGGTACCAGATTCAG	GGTAGATGACTACATATATTGACCCCTT	A	G
g47320	GAACATGTTTAAAGCTAAACGCCAT	GAACATGTTTAAAGCTAAACGCCAG	GAGCGAGTTACGGAGATAAGATCAAATAA	A	C
g47670	GAAGCATTGCTCATATCTCCAACG	CGAAGCATTGCTCATATCTCCAACA	AAACTGCTAGCCCCCCTAAGCTATA	G	A
g47720	GAGTATATGTCTATGGAATGGAAATAACA	GAGTATATGTCTATGGAATGGAAATAACG	CATAGCCATCTAAGGCAATGCTTATTGTA	T	C
g47750	ATATGGCTTTCTCAACATTGTGAACCA	ATGGCTTTCTCAACATTGTGAACCG	CTTCACATACTAGCCTTGCAAATGTCATA	T	C
g48810	TAACAGAATTAGTGTAGACTCGTGTC	GTTAACAGAATTAGTGTAGACTCGTGTA	GATGACACTATCCTGGCCATTAYTGTT	C	A
g48890	ACGATACTGGTGCCCGACGTA	CGATACTGGTGCCCGACGTG	ACAACACAGGACGCTGGATTTAYGTAT	T	C
g50450	GGGCCCGCCTTCACTAATTCTAA	GGCCCGCCTTCACTAATTCTAG	CATCTAACCCTGTTATTCAAACACCTCTT	T	C
g50670	GTAGCAATTGCAGCCGATGCAG	GGTAGCAATTGCAGCCGATGCAA	TAAGCCCCTGGATGCGCACGTT	C	T
g50680	ATGACAGGTCAAACACGATCGAGA	GACAGGTCAAACACGATCGAGC	CAAGAAGAAGCGCAAGCGATCAACTA	T	G
g54920	AATCAAGCATAACCCACCTCGACT	CAAGCATAACCCACCTCGACC	CAGGTTCCTGCTGTTGATCAACCAT	A	G
g55560	AGGAATGGTCGAAACACTTGCTTG	GAGGAATGGTCGAAACACTTGCTTT	CACATCGGATTCCCCACTTCCTTTT	C	A
g55740	TTTTATTTTGTCCCAAGCTAATTTGAGG	CTTTTTATTTTGTCCCAAGCTAATTTGAGA	GACGCGGAGTTTCTGCACCCAA	G	A
g56120	CCACAGGTAATTGCATCGCTAGAT	CACAGGTAATTGCATCGCTAGAC	GCGGATGCTCTGGCTCTTCCTT	A	G
g56130	GATGGATTTTTCAGCCCACTATATTC	GTTGATGGATTTTTCAGCCCACTATATTT	CAAGACCAAATATGTTGACTCCGAGTTT	G	A
g57740	CCGAGTTGTTTGATACTGAAGTCC	GCCGAGTTGTTTGATACTGAAGTCA	GTTCAGAGAGTTATTTGCATTTACGAGCTT	G	T
g57840	CCACTTGGTACATGTGTTAGACAATT	CCACTTGGTACATGTGTTAGACAATA	CAACCAGTCTTGCCAGCCTCATATA	A	T
g57920	GGCTTGCGAACTCGTACTTCTAC	AAGGCTTGCGAACTCGTACTTCTAT	CTCCAGGCAAGTAAGCATGTGGATT	G	A
g58150	TCTGCGTTGATGTAATGACCT	CTTCTGCGTTGATGTAATGACCC	CAGCACCCCCATCGACCAACAA	A	G
g58370	GAAAAACATGTTTTACAGCAACATTTAATAGG	AGAAAAACATGTTTTACAGCAACATTTAATAGA	ACTTGGGTGTTGTTGTCAGAGTTAGTTTA	G	A
g58570	GCATCCTATCGCAGGTGACCG	GCATCCTATCGCAGGTGACCA	CCAACATCAAACCATCAGTAGTAGTACTA	C	T
g58840	AGTGAACTTGTGACCTTGCACTCAT	GAACTTGTGACCTTGCACTCAG	GCAAACTGAAGCATTGATACCTGACAT	T	G
g58870	GGCGGCCCTGAGGTATAAATCA	GCGGCCCTGAGGTATAAATCG	TCAACGTGGAAGTGATTGATGATGATGTT	A	G
g59740	AATGCCCAGAGTTGACATTCATGC	AATAATGCCCAGAGTTGACATTCATGT	TGCATTTGATTTGCCCTGCTTTACTGAAA	G	A
g60150	TAACCTACAAAACTGTTTCTAACTTAATTTAT	AACCTACAAAACTGTTTCTAACTTAATTTAC	CTAGCTAGAAAATCCTGTAGGCTTTTGTA	A	G
g60570	TACCAAACCCAATGTTGAGCC	CTTACCAAACCCAATGTTGAGCG	ATTAGTACCAGCAGCCCTGAAGGAA	C	G
g60660	TACAATTTCCATGAGGRAAAGATAGAAAT	ACAATTTCCATGAGGRAAAGATAGAAAC	GACCATTGACCATGCATGCATGTCTT	T	C
g61070	CATGTTCTGATACAGTCCATAGGTATAA	ATGTTCTGATACAGTCCATAGGTATAC	AAACTAAGCATGGTATTCTACTGCACCTA	T	G
g62500	GGTGTAAAGACGAATCATACACATTGAA	GTGTAAAGACGAATCATACACATTGAG	TGGTGAGTTGGTGGTGGAGCAATAT	T	C
InDel_g01210	CAAATGTGGCAAAGGAAGGT	GCCGAATGAGTTCACCAGTT			
InDel_g09780	CTATGGTCTTGTCGGGGAGC	CACCAGAGGAGCGAGAGTTG			
InDel_g56500	GCATTTTCCACTGGCACGTT	AAAGCAGGAGGTCAAGAGCC			

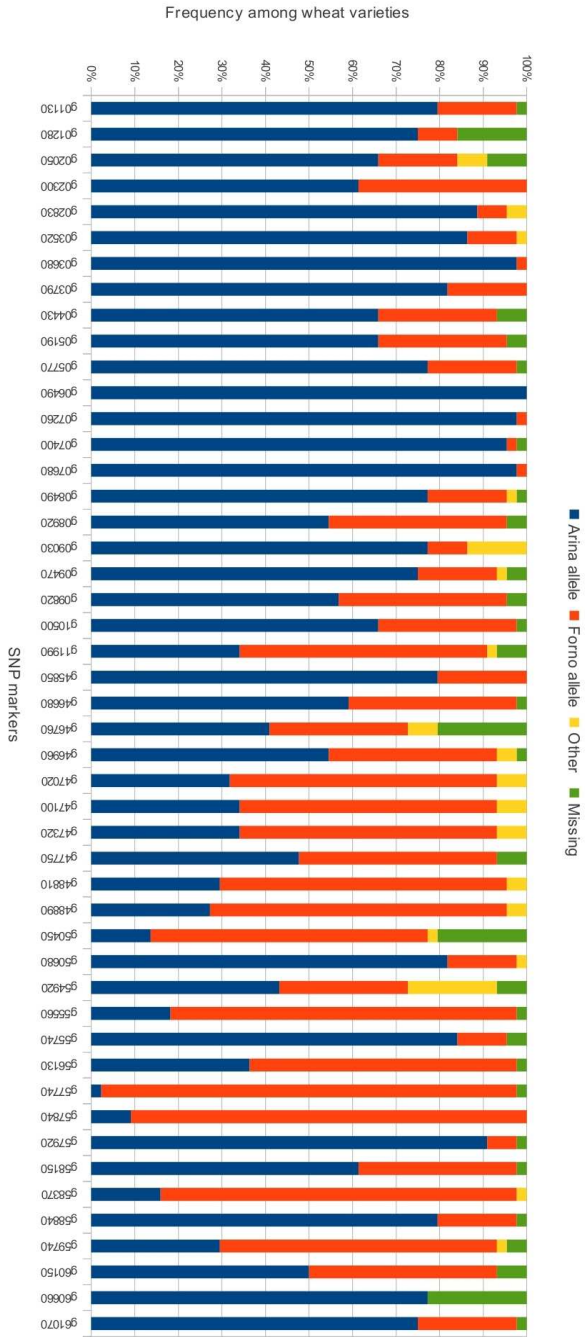

Figure S2. Arina and Forno SNP allele distribution in 44 genotyped wheat varieties.

Table S2. Genotyping results for 47 wheat varieties tested with a subset of 48 developed SNP markers

Variety	Origin	g01130	g01280	g02050	g02300	g02830	g03520	g03680	g03790	g04430	g05190	g05770	g06490	g07260	g07400	g07680	g08490	g08920	g09030	g09470	g09820	g10500	g11990	g45850	g46680
Campala	Switzerland	A:A	C:C	C:C	C:C	C:C	G:G	C:C	T:T	?	?	G:G	C:C	A:A	T:T	C:C	C:C	G:G	G:A	A:A	G:G	T:T	T:T	C:C	T:T
Fiorina	Switzerland	A:A	C:C	C:T	T:T	C:C	G:G	C:C	T:T	?	A:A	G:G	C:C	A:A	T:T	C:C	C:C	G:G	G:A	A:A	G:G	T:T	T:C	C:C	T:T
Mulan	Germany	A:A	C:C	C:T	T:T	T:T	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	C:C	C:C	C:C
Rubli	Switzerland	A:A	C:C	C:C	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	C:T	?	A:A	A:A	?	T:T	T:T	C:C	T:T
Rustic	France	A:A	C:C	C:T	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	T:T	T:T
Tapidor	France	A:A	C:C	C:T	T:T	C:C	G:G	C:C	T:T	T:T	G:G	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	C:C	C:C
Winnetou	Germany	A:A	C:C	C:T	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	?	A:A	T:T	T:C	T:T	C:C
Runal	Switzerland	A:A	C:C	?	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	G:G	A:A	G:G	G:G	C:C	T:C	C:C	C:C
Arina	Switzerland	**A:A**	**C:C**	**C:T**	**T:T**	**C:C**	**G:G**	**C:C**	**T:T**	**T:T**	**A:A**	**G:G**	**C:C**	**A:A**	**T:T**	**C:C**	**T:T**	**C:C**	**A:A**	**A:A**	**A:A**	**T:T**	**T:C**	**?**	**C:C**
Cambrena	Switzerland	A:A	C:C	C:T	C:C	T:T	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	T:T	T:T
Camedo	Switzerland	A:A	C:C	C:T	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	G:G	A:A	G:G	G:G	C:C	T:C	C:C	C:C
Caphorn	UK	A:A	C:C	C:T	T:T	C:C	G:G	C:C	T:T	T:T	G:G	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	C:C	C:C
Claro	Switzerland	A:A	C:C	C:T	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	G:G	A:A	G:G	G:G	C:C	T:C	C:C	C:C
Combin	Switzerland	A:A	?	C:T	C:C	C:C	G:G	C:C	C:C	G:G	A:A	G:G	C:C	A:A	?	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Forel	Switzerland	G:G	G:G	C:T	T:T	C:C	G:G	C:C	T:T	T:T	G:G	G:G	C:C	G:G	C:C	G:G	C:C	G:G	G:G	G:G	G:G	C:C	T:T	T:T	T:T
Galaxie	France	A:A	C:C	C:T	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	C:C	G:G	G:A	A:A	G:G	T:T	T:T	C:C	C:C
Impression	Germany	A:A	?	C:C	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	T:T	T:T
Levis	Switzerland	A:A	C:C	C:T	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Ludwig	Austria	A:A	?	C:T	C:C	C:C	C:C	T:T	C:C	G:G	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	C:C	T:T
Manhatten	Germany	A:A	C:C	C:T	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	T:T	T:T
Nara	Switzerland	A:A	C:C	C:T	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Rainer	Austria	A:A	C:C	C:C	C:C	T:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	T:T	T:T
Scaletta	Switzerland	G:G	C:C	C:T	T:T	C:C	G:G	C:C	C:C	G:G	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Segor	Switzerland	A:A	C:C	?	T:T	C:C	G:G	C:C	T:T	T:T	G:G	A:A	C:C	A:A	T:T	C:C	C:C	G:G	G:A	A:A	G:G	T:T	T:C	C:C	T:T
Sertori	Switzerland	A:A	C:C	?	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	?	G:G	G:A	?	G:G	T:T	?	C:C	T:T
Siala	Switzerland	A:A	C:C	C:T	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	G:G	A:A	G:G	G:G	C:C	T:C	C:C	?
Suretta	Switzerland	A:A	?	C:T	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Titlis	Switzerland	A:A	?	C:T	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Zinal	Switzerland	A:A	C:C	C:T	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	C:C	G:G	G:A	A:A	A:A	T:T	T:T	C:C	C:C
Forno	Switzerland	**G:G**	**G:G**	**C:C**	**C:C**	**T:T**	**C:C**	**T:T**	**C:C**	**G:G**	**G:G**	**A:A**	**T:T**	**G:G**	**C:C**	**G:G**	**C:C**	**G:G**	**G:G**	**G:G**	**G:G**	**C:C**	**T:T**	**T:T**	**T:T**
Glenlea	Canada	A:A	C:C	C:C	T:T	C:C	G:G	C:C	T:T	G:G	G:G	A:A	C:C	A:A	T:T	C:C	C:C	G:G	G:G	A:A	A:A	C:C	T:T	C:C	C:C
Pavon76	Mexico	G:G	G:G	C:T	T:T	C:C	C:C	C:C	C:C	G:G	G:G	A:A	C:C	A:A	T:T	C:C	T:T	G:G	A:A	A:A	G:G	C:C	T:T	C:C	T:T
Thatcher	Mexico	G:G	G:G	C:T	T:T	C:C	C:C	C:C	C:C	G:G	G:G	A:A	C:C	A:A	T:T	C:C	T:T	G:G	A:A	G:G	G:G	C:C	T:T	C:C	C:C
Bobwhite	Mexico	G:G	C:C	C:C	C:C	C:C	C:C	C:C	T:T	T:T	G:G	A:A	C:C	A:A	T:T	C:C	T:T	G:G	A:A	A:A	G:G	C:C	T:T	C:C	C:C
Chinese Spring	China	A:A	C:C	T:T	T:T	C:C	G:G	C:C	T:T	G:G	G:G	A:A	C:C	A:A	T:T	C:C	T:T	G:G	G:G	A:A	G:G	C:C	T:T	C:C	C:C
Chancellor	USA	G:G	C:C	C:T	C:C	C:C	G:G	C:C	T:T	G:G	A:A	A:A	C:C	A:A	T:T	C:C	C:C	G:G	G:G	G:G	G:G	C:C	T:T	C:C	C:C
Kanzler	Germany	A:A	?	C:C	C:C	T:T	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	C:C	T:T	C:C	T:T
Renan	France	G:G	C:C	?	T:T	C:C	G:G	C:C	C:C	G:G	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Amigo	USA	G:G	G:G	T:T	T:T	C:C	G:G	C:C	C:C	G:G	G:G	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:C	C:C	C:C
Disponent	Germany	A:A	C:C	C:T	T:T	T:C	C:G	C:C	T:T	T:T	?	?	C:C	A:A	T:T	C:C	T:T	?	A:A	G:A	?	?	?	C:C	C:C
MarisHuntsman	UK	A:A	C:C	C:T	C:C	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	T:T	T:T
Normandie	USA	A:A	C:C	C:T	T:T	C:C	G:G	C:C	C:C	G:G	G:G	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	C:C	C:C
TP1142XStarke	Sweden	A:A	?	C:C	T:T	C:C	G:G	C:C	T:T	T:T	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	T:T	T:T	T:T
Transec	USA	?	C:C	C:T	T:T	C:C	G:G	C:C	T:T	?	A:A	G:G	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	A:A	T:T	?	C:C	T:T
Chineselandr34	China	A:A	C:C	C:T	T:T	C:C	C:C	C:C	T:T	T:T	G:G	A:A	C:C	A:A	T:T	C:C	T:T	G:G	A:A	G:G	G:G	C:C	T:T	C:C	C:C
Chineselandr40	China	A:A	C:C	T:T	T:T	C:C	G:G	C:C	T:T	G:G	G:G	A:A	C:C	A:A	T:T	C:C	T:T	C:C	A:A	A:A	G:G	C:C	T:T	C:C	C:C

Variety	Origin	g46760	g46960	g47020	g47100	g47320	g47750	g48810	g48890	g50450	g50680	g54920	g55560	g55740	g56130	g57740	g57840	g57920	g58150	g58370	g58840	g59740	g60150	g60660	g61070
Campala	Switzerland	A:A	T:T	C:C	C:C	C:C	C:C	A:A	C:C	T:T	T:T	G:G	A:A	G:G	T:T	T:T	G:G	G:G	A:A	T:T	G:G	A:A	A:A	T:T	G:G
Fiorina	Switzerland	A:A	T:T	C:C	A:A	C:C	A:A	A:A	C:C	T:T	G:G	G:G	A:A	G:G	T:T	T:T	G:G	G:G	A:A	T:T	G:G	G:G	A:A	T:T	T:T
Mulan	Germany	C:C	A:A	C:C	A:A	C:C	A:A	A:A	C:C	T:T	G:G	G:G	A:A	G:G	T:T	T:T	G:G	G:G	A:A	T:T	G:G	A:A	A:A	T:T	T:T
Rubli	Switzerland	A:C	?	C:C	C:C	C:C	A:A	A:A	C:C	?	G:G	G:A	A:A	A:A	T:T	T:T	G:G	A:A	G:G	T:T	G:G	A:A	G:G	T:T	T:T
Rustic	France	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	C:C	G:G	G:A	A:A	A:A	T:T	T:T	G:G	A:A	G:G	T:T	G:G	A:A	G:G	T:T	G:G
Tapidor	France	?	A:A	C:C	C:C	C:C	A:A	A:A	C:C	?	T:T	G:G	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	A:A	A:A	A:A	T:T	T:T
Winnetou	Germany	C:C	T:T	C:C	C:C	C:C	A:A	A:A	C:C	C:C	T:T	G:G	C:C	G:G	T:T	T:T	G:G	A:A	A:A	T:T	A:A	A:A	G:G	?	T:T
Runal	Switzerland	C:C	A:A	T:T	A:A	T:T	C:C	T:T	T:T	T:T	T:T	A:A	A:A	G:G	T:T	T:T	G:G	G:G	A:A	T:T	G:G	G:G	A:A	T:T	T:T
Arina	**Switzerland**	**C:C**	**A:A**	**T:T**	**A:A**	**T:T**	**C:C**	**C:C**	**T:T**	**T:T**	**T:T**	**G:A**	**C:C**	**G:G**	**A:A**	**G:G**	**A:A**	**G:G**	**A:A**	**T:T**	**G:G**	**A:A**	**T:T**	**T:T**	**T:T**
Cambrena	Switzerland	A:A	T:T	C:C	C:C	C:C	C:C	C:C	C:C	C:C	T:T	G:G	A:A	G:G	A:A	T:T	G:G	A:A	A:A	T:T	T:T	A:A	G:G	T:T	G:G
Cameo	Switzerland	C:C	A:A	C:C	A:A	T:T	A:A	A:A	C:C	T:T	T:T	A:A	A:A	G:G	T:T	T:T	G:G	A:A	G:G	T:T	T:T	G:G	A:A	T:T	T:T
Caphorn	UK	?	A:A	C:C	C:C	T:T	C:C	A:A	C:C	?	T:T	G:A	A:A	G:G	T:T	T:T	G:G	A:A	G:G	T:T	T:T	A:A	G:G	?	T:T
Claro	Switzerland	C:C	A:A	A:A	T:T	T:T	C:C	A:A	C:C	C:C	T:T	A:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	T:T	G:G	A:A	?	T:T
Combin	Switzerland	C:C	A:A	A:A	A:A	T:T	C:C	A:A	C:C	C:C	G:G	G:A	C:C	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	?	T:T
Forel	Switzerland	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	?	G:G	G:G	A:A	A:A	T:T	T:T	A:A	G:G	A:A	T:T	G:G	A:A	G:G	?	G:G
Galaxie	France	?	A:A	C:C	C:C	T:T	C:C	A:A	C:C	?	T:T	G:G	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	?	T:T
Impression	Germany	A:A	T:T	C:C	A:A	C:C	A:A	A:A	C:C	T:T	G:G	G:G	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Levis	Switzerland	C:C	A:A	T:T	A:A	T:T	C:C	A:A	C:C	T:T	T:T	G:A	A:A	G:G	T:T	T:T	G:G	A:A	A:G	T:T	G:G	?	G:G	T:T	G:G
Ludwig	Austria	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	T:T	G:G	G:A	A:A	A:A	G:G	T:T	G:G	?	A:A	T:T	G:G	A:A	A:A	T:T	G:G
Manhatten	Germany	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	T:T	G:G	G:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Nara	Switzerland	C:C	A:A	C:C	A:A	C:C	C:C	C:C	T:T	T:T	G:A	A:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	A:A	A:A	G:G	T:T	T:T
Rainer	Austria	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	C:C	T:T	G:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	A:A	A:A	G:G	T:T	T:T
Scaletta	Switzerland	C:C	A:A	A:A	T:T	T:T	C:C	A:A	C:C	C:C	T:T	G:G	C:C	G:G	T:T	T:T	G:G	A:A	A:A	T:T	A:A	A:A	G:G	?	G:G
Segor	Switzerland	A:A	T:T	C:C	A:A	C:C	A:C	A:A	C:T	T:T	G:A	G:G	C:C	G:G	T:T	T:T	G:G	A:A	A:A	T:T	A:A	A:A	G:G	T:T	G:G
Sentori	Switzerland	A:C	T:A	C:A	C:T	?	A:C	C:C	C:T	?	T:T	A:A	A:A	?	T:T	T:T	G:G	G:G	A:A	?	G:G	A:A	A:A	T:T	T:T
Siala	Switzerland	C:C	A:A	A:A	A:A	T:T	C:C	A:A	T:T	C:C	T:T	A:A	A:A	G:G	?	T:T	G:G	A:A	A:A	T:T	T:T	A:A	A:A	T:T	T:T
Sunetta	Switzerland	C:C	A:A	A:A	T:T	T:T	C:C	A:A	T:T	T:T	T:T	G:A	C:C	G:G	T:T	T:T	G:G	A:A	G:G	T:T	T:T	G:G	A:A	T:T	T:T
Tillis	Switzerland	C:C	A:A	A:A	A:A	T:T	C:C	A:A	T:T	T:T	T:T	G:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	T:T	A:A	G:G	T:T	G:G
Zinal	Switzerland	?	A:A	C:C	C:C	T:T	A:A	A:A	C:C	?	T:T	G:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	A:A	A:A	A:A	T:T	T:T
Forno	**Switzerland**	**A:A**	**T:T**	**C:C**	**C:C**	**C:C**	**C:C**	**A:A**	**C:C**	**C:C**	**G:G**	**G:G**	**A:A**	**A:A**	**G:G**	**T:T**	**A:A**	**G:G**	**A:A**	**G:G**	**G:G**	**A:A**	**G:G**	**T:T**	**G:G**
Glenlea	Canada	A:A	T:T	C:C	C:C	C:C	C:C	A:A	C:C	C:C	G:G	A:A	A:A	G:G	T:T	T:T	A:A	G:G	A:A	T:T	G:G	G:G	A:A	T:T	G:G
Pavon76	Mexico	A:A	T:T	C:C	C:C	C:C	C:C	A:A	C:C	C:C	T:T	G:A	A:A	G:G	T:T	T:T	G:G	G:G	A:A	T:T	G:G	G:G	A:A	T:T	G:G
Thatcher	Mexico	?	A:A	C:C	C:C	T:T	C:C	A:A	C:C	?	T:T	A:A	A:A	?	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	A:A	T:T	T:T
Bobwhite	Mexico	?	A:A	C:C	A:A	T:T	C:C	A:A	C:C	?	T:T	A:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Chinese Spring	China	C:C	T:T	A:A	T:T	?	?	A:A	C:T	C:C	T:T	A:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Chancellor	USA	?	A:A	C:C	C:C	C:C	A:C	A:A	C:T	C:C	T:T	G:G	A:A	G:G	T:T	T:T	A:A	G:G	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Kanzler	Germany	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	C:C	G:G	G:A	A:A	G:G	T:T	T:T	A:A	G:G	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Renan	France	C:C	A:A	C:C	A:A	C:C	C:C	A:A	C:C	C:C	T:T	G:A	C:C	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	?	T:T	G:G
Amigo	USA	C:C	A:A	C:C	A:A	C:C	A:A	A:A	C:C	C:C	T:T	G:A	C:C	G:G	T:T	T:T	?	G:G	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Disponent	Germany	C:C	A:A	C:A	C:T	T:T	A:A	A:A	C:C	C:C	G:T	A:A	?	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	T:T
MarisHuntsman	UK	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	C:C	G:G	G:G	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	T:T
Normandie	USA	C:C	A:A	C:C	C:C	C:C	A:A	A:A	C:C	C:C	T:T	G:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	T:T
TP1142XStarke	Sweden	A:A	T:T	C:C	C:C	?	A:A	A:A	C:C	C:C	T:T	G:A	C:C	G:G	T:T	T:T	A:A	A:A	A:A	T:T	A:G	A:A	A:A	T:T	T:T
Transec	USA	A:C	T:A	C:A	C:T	?	A:C	A:A	C:T	C:C	T:T	?	C:C	G:G	T:T	T:T	G:G	A:A	G:G	T:T	A:G	A:A	A:A	T:T	T:T
Chineseland34	China	A:A	T:T	C:C	C:C	C:C	A:A	A:A	C:C	?	T:T	G:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	G:G	T:T	G:G
Chineseland40	China	?	T:T	C:C	C:C	C:C	A:A	A:A	C:C	C:C	T:T	A:A	A:A	G:G	T:T	T:T	G:G	A:A	A:A	T:T	G:G	A:A	A:A	T:T	G:G

Table S3. Ratio of detected genes with SNP to the total number of genes calculated using sliding windows

Window Start	Window end	N genes with SNP	Total N of genes	SNP/total genes ratio
0	1200000	22	107	0.206
100000	1300000	25	113	0.221
200000	1400000	26	125	0.208
300000	1500000	25	122	0.205
400000	1600000	24	126	0.190
500000	1700000	21	124	0.169
600000	1800000	20	120	0.167
700000	1900000	19	112	0.170
800000	2000000	16	111	0.144
900000	2100000	14	100	0.140
1000000	2200000	16	112	0.143
1100000	2300000	16	110	0.145
1200000	2400000	16	110	0.145
1300000	2500000	10	102	0.098
1400000	2600000	10	100	0.100
1500000	2700000	13	117	0.111
1600000	2800000	15	123	0.122
1700000	2900000	16	125	0.128
1800000	3000000	15	129	0.116
1900000	3100000	16	133	0.120
2000000	3200000	18	141	0.128
2100000	3300000	18	145	0.124
2200000	3400000	16	138	0.116
2300000	3500000	16	137	0.117
2400000	3600000	16	144	0.111
2500000	3700000	16	135	0.119
2600000	3800000	16	128	0.125
2700000	3900000	14	114	0.123
2800000	4000000	13	110	0.118
2900000	4100000	12	104	0.115
3000000	4200000	12	102	0.118
3100000	4300000	11	100	0.110
3200000	4400000	11	89	0.124
3300000	4500000	11	93	0.118
3400000	4600000	11	89	0.124
3500000	4700000	11	92	0.120
3600000	4800000	11	88	0.125
3700000	4900000	11	88	0.125
3800000	5000000	13	95	0.137
3900000	5100000	11	94	0.117
4000000	5200000	9	89	0.101
4100000	5300000	8	88	0.091
4200000	5400000	9	94	0.096
4300000	5500000	8	96	0.083
4400000	5600000	8	96	0.083
4500000	5700000	8	95	0.084
4600000	5800000	10	103	0.097
4700000	5900000	10	101	0.099
4800000	6000000	11	105	0.105
4900000	6100000	14	112	0.125
5000000	6200000	11	106	0.104
5100000	6300000	11	100	0.110
5200000	6400000	12	99	0.121
5300000	6500000	13	102	0.127
5400000	6600000	14	95	0.147
5500000	6700000	15	96	0.156
5600000	6800000	13	94	0.138
5700000	6900000	16	95	0.168
5800000	7000000	14	91	0.154
5900000	7100000	13	87	0.149
6000000	7200000	15	89	0.169
6100000	7300000	15	89	0.169

6200000	7400000	16	89	0.180
6300000	7500000	16	95	0.168
6400000	7600000	16	95	0.168
6500000	7700000	21	101	0.208
6600000	7800000	20	103	0.194
6700000	7900000	20	99	0.202
6800000	8000000	23	106	0.217
6900000	8100000	24	112	0.214
7000000	8200000	24	111	0.216
7100000	8300000	24	114	0.211
7200000	8400000	21	108	0.194
7300000	8500000	19	108	0.176
7400000	8600000	18	101	0.178
7500000	8700000	19	102	0.186
7600000	8800000	19	102	0.186
7700000	8900000	13	91	0.143
7800000	9000000	12	88	0.136
7900000	9100000	12	92	0.130
8000000	9200000	10	85	0.118
8100000	9300000	7	78	0.090
8200000	9400000	7	72	0.097
8300000	9500000	7	65	0.108
8400000	9600000	7	60	0.117
8500000	9700000	6	52	0.115
8600000	9800000	8	62	0.129
8700000	9900000	7	53	0.132
8800000	10000000	6	48	0.125
8900000	10100000	6	44	0.136
9000000	10200000	6	40	0.150
9100000	10300000	6	33	0.182
9200000	10400000	6	30	0.200
9300000	10500000	6	31	0.194
9400000	10600000	6	32	0.188
9500000	10700000	6	33	0.182
9600000	10800000	6	36	0.167
9700000	10900000	6	39	0.154
9800000	11000000	4	30	0.133
9900000	11100000	4	32	0.125
10000000	11200000	5	37	0.135
10100000	11300000	6	40	0.150
10200000	11400000	7	42	0.167
10300000	11500000	6	43	0.140
10400000	11600000	5	41	0.122
10500000	11700000	3	35	0.086
10600000	11800000	4	37	0.108
10700000	11900000	6	41	0.146
10800000	12000000	6	39	0.154
10900000	12100000	6	37	0.162
11000000	12200000	6	41	0.146
11100000	12300000	6	41	0.146
11200000	12400000	5	37	0.135
11300000	12500000	4	38	0.105
11400000	12600000	3	37	0.081
11500000	12700000	3	35	0.086
11600000	12800000	3	34	0.088
11700000	12900000	3	31	0.097
11800000	13000000	2	31	0.065
39200000	40400000	1	10	0.100
39300000	40500000	1	11	0.091
39400000	40600000	1	13	0.077
39500000	40700000	1	17	0.059
39600000	40800000	1	18	0.056
39700000	40900000	1	19	0.053
39800000	41000000	1	25	0.040
39900000	41100000	1	31	0.032

40000000	41200000	1	34	0.029
40100000	41300000	2	35	0.057
40200000	41400000	2	40	0.050
40300000	41500000	3	51	0.059
40400000	41600000	3	53	0.057
40500000	41700000	3	56	0.054
40600000	41800000	4	61	0.066
40700000	41900000	4	60	0.067
40800000	42000000	5	63	0.079
40900000	42100000	5	66	0.076
41000000	42200000	5	64	0.078
41100000	42300000	5	60	0.083
41200000	42400000	5	62	0.081
41300000	42500000	4	62	0.065
41400000	42600000	4	63	0.063
41500000	42700000	4	59	0.068
41600000	42800000	4	62	0.065
41700000	42900000	4	60	0.067
41800000	43000000	4	60	0.067
41900000	43100000	5	65	0.077
42000000	43200000	4	68	0.059
42100000	43300000	5	71	0.070
42200000	43400000	5	69	0.072
42300000	43500000	6	76	0.079
42400000	43600000	7	77	0.091
42500000	43700000	8	79	0.101
42600000	43800000	8	77	0.104
42700000	43900000	8	79	0.101
42800000	44000000	8	77	0.104
42900000	44100000	10	84	0.119
43000000	44200000	9	83	0.108
43100000	44300000	9	83	0.108
43200000	44400000	10	83	0.120
43300000	44500000	10	81	0.123
43400000	44600000	11	89	0.124
43500000	44700000	11	84	0.131
43600000	44800000	10	78	0.128
43700000	44900000	10	78	0.128
43800000	45000000	10	80	0.125
43900000	45100000	9	74	0.122
44000000	45200000	8	72	0.111
44100000	45300000	6	68	0.088
44200000	45400000	6	68	0.088
44300000	45500000	5	69	0.072
44400000	45600000	4	68	0.059
44500000	45700000	3	71	0.042
44600000	45800000	3	71	0.042
44700000	45900000	3	75	0.040
44800000	46000000	3	77	0.039
44900000	46100000	2	76	0.026
45000000	46200000	2	77	0.026
45100000	46300000	3	82	0.037
45200000	46400000	4	88	0.045
45300000	46500000	4	89	0.045
45400000	46600000	4	88	0.045
45500000	46700000	5	91	0.055
45600000	46800000	6	89	0.067
45700000	46900000	6	89	0.067
45800000	47000000	5	89	0.056
45900000	47100000	5	89	0.056
46000000	47200000	8	97	0.082
46100000	47300000	11	106	0.104
46200000	47400000	14	109	0.128
46300000	47500000	17	115	0.148
46400000	47600000	18	112	0.161

46500000	47700000	18	115	0.157
46600000	47800000	18	117	0.154
46700000	47900000	17	106	0.160
46800000	48000000	16	110	0.145
46900000	48100000	18	111	0.162
47000000	48200000	19	110	0.173
47100000	48300000	18	111	0.162
47200000	48400000	16	111	0.144
47300000	48500000	13	107	0.121
47400000	48600000	10	104	0.096
47500000	48700000	6	98	0.061
47600000	48800000	5	98	0.051
47700000	48900000	6	101	0.059
47800000	49000000	6	105	0.057
47900000	49100000	8	118	0.068
48000000	49200000	8	118	0.068
48100000	49300000	6	117	0.051
48200000	49400000	5	115	0.043
48300000	49500000	5	112	0.045
48400000	49600000	4	110	0.036
48500000	49700000	4	109	0.037
48600000	49800000	4	110	0.036
48700000	49900000	4	113	0.035
48800000	50000000	4	117	0.034
48900000	50100000	3	115	0.026
49000000	50200000	3	112	0.027
49100000	50300000	1	106	0.009
49200000	50400000	3	108	0.028
49300000	50500000	5	113	0.044
49400000	50600000	6	122	0.049
49500000	50700000	6	127	0.047
49600000	50800000	6	129	0.047
49700000	50900000	6	127	0.047
49800000	51000000	7	127	0.055
49900000	51100000	7	126	0.056
50000000	51200000	7	124	0.056
50100000	51300000	8	123	0.065
50200000	51400000	10	124	0.081
50300000	51500000	10	131	0.076
50400000	51600000	10	130	0.077
50500000	51700000	8	128	0.063
50600000	51800000	8	120	0.067
50700000	51900000	8	123	0.065
50800000	52000000	10	122	0.082
50900000	52100000	10	127	0.079
51000000	52200000	11	131	0.084
51100000	52300000	11	128	0.086
51200000	52400000	11	131	0.084
51300000	52500000	11	136	0.081
51400000	52600000	9	138	0.065
51500000	52700000	9	128	0.070
51600000	52800000	8	129	0.062
51700000	52900000	8	131	0.061
51800000	53000000	7	137	0.051
51900000	53100000	8	135	0.059
52000000	53200000	7	134	0.052
52100000	53300000	7	127	0.055
52200000	53400000	5	125	0.040
52300000	53500000	5	129	0.039
52400000	53600000	5	128	0.039
52500000	53700000	5	125	0.040
52600000	53800000	6	126	0.048
52700000	53900000	8	137	0.058
52800000	54000000	7	137	0.051
52900000	54100000	7	134	0.052

53000000	54200000	8	132	0.061
53100000	54300000	8	132	0.061
53200000	54400000	7	123	0.057
53300000	54500000	9	130	0.069
53400000	54600000	11	128	0.086
53500000	54700000	11	129	0.085
53600000	54800000	10	125	0.080
53700000	54900000	9	121	0.074
53800000	55000000	8	120	0.067
53900000	55100000	7	112	0.063
54000000	55200000	7	112	0.063
54100000	55300000	7	106	0.066
54200000	55400000	6	102	0.059
54300000	55500000	6	103	0.058
54400000	55600000	6	110	0.055
54500000	55700000	5	112	0.045
54600000	55800000	3	114	0.026
54700000	55900000	4	110	0.036
54800000	56000000	8	117	0.068
54900000	56100000	9	121	0.074
55000000	56200000	12	126	0.095
55100000	56300000	13	129	0.101
55200000	56400000	15	133	0.113
55300000	56500000	17	140	0.121
55400000	56600000	21	142	0.148
55500000	56700000	23	139	0.165
55600000	56800000	25	145	0.172
55700000	56900000	24	134	0.179
55800000	57000000	25	131	0.191
55900000	57100000	24	135	0.178
56000000	57200000	21	132	0.159
56100000	57300000	20	130	0.154
56200000	57400000	21	125	0.168
56300000	57500000	20	129	0.155
56400000	57600000	22	135	0.163
56500000	57700000	20	130	0.154
56600000	57800000	18	133	0.135
56700000	57900000	16	128	0.125
56800000	58000000	19	135	0.141
56900000	58100000	22	142	0.155
57000000	58200000	25	146	0.171
57100000	58300000	28	144	0.194
57200000	58400000	28	137	0.204
57300000	58500000	29	138	0.210
57400000	58600000	26	141	0.184
57500000	58700000	25	139	0.180
57600000	58800000	25	136	0.184
57700000	58900000	25	142	0.176
57800000	59000000	25	144	0.174
57900000	59100000	26	152	0.171
58000000	59200000	23	145	0.159
58100000	59300000	21	147	0.143
58200000	59400000	17	136	0.125
58300000	59500000	14	126	0.111
58400000	59600000	13	121	0.107
58500000	59700000	12	111	0.108
58600000	59800000	11	97	0.113
58700000	59900000	11	84	0.131

CHAPTER 3

High-resolution analysis of a QTL for resistance to Stagonospora nodorum glume blotch in wheat reveals presence of two distinct resistance loci in the target interval

Margarita Shatalina, Monika Messmer, Catherine Feuillet, Fabio Mascher, Etienne Paux, Frédéric Choulet, Thomas Wicker and Beat Keller

in preparation

Abstract

Stagonospora nodorum glume blotch (SNG), caused by the necrotrophic fungus *Stagonospora nodorum,* is one of the economically important diseases of bread wheat (*Triticum aestivum* L.). Resistance to SNG is known to be quantitative and previous studies of a recombinant inbred line (RIL) population identified a major quantitative trait locus (QTL) for resistance to SNG on the short arm of chromosome 3B. To localize this QTL (*QSng.sfr-3BS*) with high resolution, we constructed a genetic map for the QTL target region using information from sequenced flow-sorted chromosomes 3B of the two parental cultivars, the physical map of chromosome 3B of cultivar 'Chinese Spring' and BAC-clone sequences. The mapping population of near-isogenic lines (NIL) was evaluated for SNG resistance in field infection tests. NILs segregated for disease resistance as well as for plant height, additionally, we observed a high environmental influence on the trait. Our analysis detected a strong negative correlation of SNG resistance and plant height. Further analysis of the target region identified two linked loci associated with SNG resistance. One of them was also associated with plant height, revealing an effect of *QSng.sfr-3BS* on plant height that was hidden in the RIL population. This result demonstrates an unexpectedly high genetic complexity of resistance controlled by *QSng.sfr-3BS* and shows the importance of the study of QTL in mendelized form in NILs.

Introduction

Bread wheat (*Triticum aestivum* L.) is globally one of the most important crops in terms of cultivated area and food source. In 2010, FAO reported a production of approximately 654 million tons of wheat worldwide (FAO stat 2010). *Phaeosphaeria nodorum* (E. Müller) Hedjaroude (anamorph *Stagonospora nodorum* (Berk.) Castellani and Germano) is a necrotrophic fungus which infects both wheat glumes, causing Stagonospora nodorum glume blotch (SNG), and leaves, causing Stagonospora nodorum leaf blotch (SNL). SNG and SNL lead to substantial yield losses in many wheat growing areas, including Europe, North America and Australia (Duczek *et al.* 1999; Halama 2002; Oliver *et al.* 2012). It was shown that the infection progress highly depends on the environmental conditions. For example, the analysis of the large dataset from Rothamsted Broadbalk experiment archive, which included the records from 1844 to 2003, suggested that fluctuations in *S. nodorum* epidemics depend on the amount of rainfalls, temperature and SO_2 emissions (Shaw *et al.* 2008). The importance of wetness (relative humidity and precipitations) and temperature for the SNG progress were demonstrated in field and growth chamber experiments (Wolf and Francl 2000; Shah and Bergstrom 2002).

Resistance to SNG and SNL in bread wheat have been shown to be genetically distinct (Laubscher *et al.* 1966; Broennimann 1975; Nelson and Gates 1982; Fried and Meister 1987; Bostwick *et al.* 1993; Van Ginkel and Rajaram 1999; Wicki *et al.* 1999). To date, only quantitative resistance to SNG has been reported. It has been identified in several wheat populations independently (Schnurbusch *et al.* 2003; Aguilar *et al.* 2005). Schnurbusch *et al.* (2003) performed a QTL mapping study on a population of 240 recombinant inbred lines (RIL) developed from a cross of cultivars 'Arina' and 'Forno'. This study identified two important QTLs providing resistance to SNG: a major QTL, *QSng.sfr-3BS,* was detected on the short arm of chromosome 3B and explained 31.2% of the phenotypic variance for the disease resistance within the population. The second QTL *QSng.sfr-4BL*, explaining 19.1% of phenotypic variance, was found on the long arm of chromosome 4B. Both detected QTLs were not co-localized with any QTLs for plant height. Additionally, an association mapping study on 44 wheat cultivars of different origin showed a significant association between the genetic marker *sun2* from the target region for *QSng.sfr-3BS* and the resistance to SNG (Tommasini *et al.* 2007).

It has been shown that plant height is often associated with resistance to splashed dispersed fungal pathogens (Eriksen *et al.* 2003; Draeger *et al.* 2007; Srinivasachary *et al.* 2009). Shorter plants show increased susceptibility to the pathogens and taller plants tend to be more resistant. There are several explanations of this negative correlation. First, in the case of a natural infection, since it progresses from the bottom to the top of the plant, taller plants get an advantage of having larger distance between the leaves (Mesterhazy *et al.* 1995). Second, it was suggested that microclimate conditions for infection development

on tall plants are often less favourable in comparison to the short plants (Scott *et al.* 1985). Third, the genes influencing plant height and disease resistance might be genetically linked (Gervais *et al.* 2003; Schnurbusch *et al.* 2003). Besides plant height, SNG resistance was shown to be associated with morphological traits such as flowering time (Broennimann *et al.* 1973; Scott *et al.* 1982; Van Ginkel and Rajaram 1999; Wicki *et al.* 1999).

Genetic mapping in bread wheat is challenging as wheat has a large (17Gb), hexaploid and highly repetitive (80-90%) genome (Wanjugi *et al.* 2009; Bennett and Smith 1976; Hollister and Gaut 2009). Furthermore, low level of polymorphism between elite wheat cultivars (Bryan *et al.* 1999) hinders the development of genetic markers. Recent advances in wheat genomics, such as reducing complexity by sorting individual chromosome arms using flow cytometry, building chromosome-specific BAC libraries and physical maps, partial genome sequencing, and adopting diploid relatives as models for characterization of the genome, contributed to the easier and faster identification of polymorphisms. In addition, they revealed that, at a high-resolution level, synteny breaks, inversions and interspersed nonsyntenic genes are frequent (Choulet *et al.* 2010), thus highlighting the limitations of using model genomes for map-based cloning in wheat. The isolation of target genes from wheat chromosome 3B benefits from the availability of a physical map (Paux *et al.* 2008 ; Rustenholz *et al.* 2011) that was developed using fingerprints from a bacterial artificial chromosomes (BAC) library of the wheat landrace 'Chinese Spring'. More than 4000 markers including simple sequence repeats (SSR) and insertion site based polymorphisms (ISBP) markers derived from BAC-end sequences have been anchored to the 3B physical map and can now be used in various mapping projects. In addition, the anchoring of numerous genes (Rustenholz *et al.* 2011) on the physical map enables to use the synteny between the regions in wheat chromosome 3B and corresponding regions on rice chromosome 1 (*Oryza sativa* L.) (The International Rice Genome Sequencing Project 2005).

Here, we report on the evaluation of a mapping population of near-isogenic lines for SNG resistance in the field which revealed a segregation for both disease resistance and plant height. Our analysis detected a negative correlation between SNG resistance and plant height. Two genetic regions associated with SNG resistance were found in the previously identified target region. One of them was associated with both, plant height and SNG resistance, suggesting that the same QTL controls both traits.

Material and methods

Plant material

The mapping population was developed from a cross between the cultivars 'Arina' and 'Forno', two Swiss winter bread wheat cultivars with contrasting SNG phenotypes: 'Arina' shows high levels of resistance to *S. nodorum* glume blotch (SNG) and is a tall variety (120 cm) whereas 'Forno' is highly susceptible to SNG and is shorter (105 cm).

F1 progeny of this cross was selfed and back-crossed to the susceptible parent 'Forno' twice. The resulting BC2 plants were tested using markers *barc133 and gwm389,* flanking the target interval of *QSng.sfr-3BS* (Schnurbusch *et al.* 2003; Tommasini *et al.* 2007), and markers *barc75* and *sun2* located in the interval. Additionally, BC2 plants were also selected for the presence of the 'Forno' allele of the second QTL from the chromosome 4B with the marker *gwm251*. The selected BC2 line was back-crossed again to 'Forno' and the resulting BC3F2 plants were screened for recombinants in the target interval between the flanking markers. The selected near-isogenic lines (NILs) were selfed and multiplied in the greenhouse. BC3F5 and BC3F6 plants of these NILs were used in the field tests.

Construction of the genetic map

We mapped in the NIL population microsatellite markers *gwm1034, barc75* and *sun2* which were previously mapped in a population of recombinant inbred lines from a cross of 'Arina' x 'Forno' (Tommasini *et al.* 2007). Genetic markers (SSR, ISBP) from the 3B physical map were amplified according to the conditions published by Paux *et al.* (2008). Additionally, new SSR markers were developed based on the sequence information from ctg344 (Choulet *et al.* 2010) which is located in the QTL interval. PCR products from SSR markers were analyzed on an ABI 3730 sequencer (Applied Biosystems, USA). Microsatellite data was analysed with the GeneMapper 4.0 software for polymorphisms (Applied Biosystems, USA). ISBP markers were detected using both gel-electrophoresis and high-resolution melting curve analysis on a CFX96™ Real Time PCR System (Bio-Rad Laboratories, Hercules, USA). Six BAC clones (TaaCsp3BFhA_0035M17, TaaCsp3BFhA_0112H11, TaaCsp3BFhA_0078F20, TaaCsp3BFhA_0027A18, TaaCsp3BFhA_0163M03, TaaCsp3BFhA_0091M22) were ordered from INRA CNRGV (Toulouse, France) and were sequenced with 454 Titanium technology by GATC (Konstanz, Germany). Newly developed BAC sequence-based InDel and CAPS markers were amplified and analyzed by gel-electrophoresis. Flow-sorted chromosomes 3B of 'Arina' and 'Forno' were prepared and sequenced as described in Shatalina *et al.* (2013). New SNP and InDel markers derived from the sequence analysis were amplified using the following cycling conditions: 10 second denaturation at 95°, annealing during 30 seconds

at 61° and 30 seconds extension at 72°, for 35 cycles. The samples were initially denaturated for 3 minutes and the cycling was followed by an extension step of 7 min. We used gel-electrophoresis to detect the amplified fragments for genotyping. SNP marker *swm01810bd* was detected using HRM on a CFX96™ Real Time PCR System (Bio-Rad Laboratories, Hercules, USA).

Disease phenotyping

Field trials with the NIL population were conducted in 2011 and 2012 at three different locations in Switzerland: Agroscope Reckenholz-Tänikon Research Station, near Zurich (environment Reckenholz); Agroscope Changins-Wädenswil Research Station, close to Nyon (environment Changins); Vouvry in canton Valais, near Lake Geneva (environment Vouvry). In Zurich, the trial was located at 443 m above sea level on loamy soils with an average precipitation of 1,000 mm and an average temperature of 7.9°C. The Vouvry site is situated at about 381 m above sea level on loamy sandy soils with an average precipitation of 881 mm and an average temperature of 8.7°C. The Changins site is located at 430 m above sea level on well-drained brown soil (calcaric cambisol) with an average precipitation of 654 mm and a mean temperature of 9.5°C. Each NIL was tested in two replications in every environment. In 2011, 92 homozygous NILs were tested in Reckenholz and, because of the insufficient number of seeds from some NILs, only 67 lines were evaluated for resistance to SNG in Changins and Vouvry environments (Table 1).

Table 1. Field assessments of SNG resistance and plant height in the NIL mapping population

Year	Environment	AUDPC	Plant height
2011	Changins	67 NILs	67 NILs
2011	Vouvry	67 NILs	-
2011	Reckenholz	-	92 NILs
2012	Reckenholz	28 NILs	28 NILs

In 2012, due to the limited resources, a subset of 28 NILs was tested in all three locations. Artificial infection was done using a motor sprayer. The infection cocktail contained a mixture of *S. nodorum* isolates from the collection of Agroscope Changins-Wädenswil. Plants were infected twice: a first infection was made at the stage before the emergence of spikes with 5 million spores/ml in 300 l water/ha; and a second infection was made at the flowering stage with 1 million spores/ml in 300 l water/ha. Disease severity was scored according to the Broennimann scale (Broennimann, 1968). This method allows to score the percentage of infected glume area and has a scale of 0, 5, 10, 25, 50, 75 and 100%.

For each environment disease development and severity was scored at least four times during its progress. Final scores were transformed to Area Under the Disease Progress Curve (AUDPC) as described in Jeger and Viljanen-Rollinson (2001). Plant height (cm) was measured in environments Reckenholz and Changins in 2011 and Reckenholz-2012 for both replications of NILs.

Statistical analysis of the phenotypic assessments

Broad-sense heritability of the traits was estimated according to Hallauer and Miranda (1981): $h^2 = \sigma^2_g / (\sigma^2_e/re + \sigma^2_{ge}/e + \sigma^2_g)$

where σ^2_g – genotypic variance, σ^2_e – environmental variance, σ^2_{ge} – variance of genotype/environment interaction, e – number of environments, r – number of replicas.

The variances were estimated from the mixed linear model:

$$Y = X\beta + Zu + Tv + \varepsilon$$

where Y is a vector of phenotypic observations, β is a vector of fixed genotype effects, u is a vector of random environmental effects, v – a vector of random effects from interaction of genotype and environment and ε – vector of residual errors. X, Z and T are coefficients of corresponding fixed and random effects. Mixed linear model was analysed using JMP 9.0 software (SAS institute, Cary, USA).

Single marker analysis was conducted in the R environment for statistical computing (R development core team, 2005) using functions 'lme' and 'lmer' following the mixed linear model:

$$Y = X\beta + Zu + Tv + \varepsilon$$

Y is a vector of phenotypic data (SNG resistance or plant height), β is a vector of fixed marker effects, u is a vector of random environmental effects, v – a vector of random effects from interaction of marker and environment and ε – vector of residual errors. X, Z and T are coefficients of corresponding fixed and random effects. Following the guidelines for model optimization (Crawley, 2007) we eliminated marker x environment interaction as non-significant factor for both AUDPC and plant height. The calculation was done using the reduced model:

$$Y = X\beta + Zu + \varepsilon$$

Plant height was included as a fixed cofactor in the general linear model ('glm' function in the R environment) to test its effect on AUDPC:

$$Y = X\beta + Zu + \varepsilon$$

Y is a vector of phenotypic data (SNG resistance), β is a vector of fixed marker effects, u is a vector of fixed effects from plant height and ε – vector of residual errors. X, Z are coefficients of corresponding fixed and random effects.

Calculations of Spearman correlation coefficient (r) for the traits between different environments and replicas was done using JMP 9.0 (SAS institute, Cary, USA). Percentage of explained variance for environment and plant height was estimated using function 'ICC1' of package 'multilevel' in the R environment for statistical analysis. *t*-tests were conducted in R.

Results

High resolution mapping of the *QSng.sfr-3BS* target region supported by the 3B physical map and whole chromosome sequencing

Schnurbusch *et al.* (2003) identified *QSng.sfr-3BS* as the major QTL for resistance to SNG in the wheat cultivar 'Arina'. This QTL is located in the telomeric part of chromosome 3B in the interval flanked by the SSR markers *gwm389* and *barc133*. To further characterize this target interval and the QTL, we developed a high-resolution mapping population consisting of NILs derived from a cross between the susceptible cultivar 'Forno' as recurrent parent and the resistant cultivar 'Arina' as donor. Among 1320 NILs (BC3F2), we selected 145 lines showing recombination between the flanking markers of the target interval. Ninety-two of them were homozygous for the flanking markers. The size of the target region for the *QSng.sfr-3BS* in our NIL population was 11 cM.

Several approaches were used to construct a high density genetic map in the target interval. In a first step, we mapped markers *gwm1034, barc75* and *sun2* from the previous map of the 'Arina' x 'Forno' RIL population (Paillard *et al.* 2003; Tommasini *et al.* 2007). We also used the availability of the chromosome 3B physical map for cultivar 'Chinese Spring' (Paux *et al.* 2008) which allowed us to estimate the physical size of the region of interest as well as to map additional markers developed for anchoring the physical map. According to the physical map, the 11 cM target interval spanned a region of approximately 10 Mbp on chromosome 3B (Figure 1). We mapped a set of SSR and ISBP markers, originally derived by Paux *et al.* (2008) for the anchoring of the physical map of chromosome 3B. Unfortunately, a vast majority of the tested ISBP markers from the physical map were not polymorphic in the 'Arina' x 'Forno' NIL population. From a total of 33 tested markers only *cfp3062* was polymorphic and was successfully added to the map. In addition, two new SSR markers (*swm105, swm106*) we designed based on the sequence information from the sequenced physical contig 344 of chromosome 3B (Choulet *et al.* 2010). These markers were polymorphic and genetically mapped to our target interval.

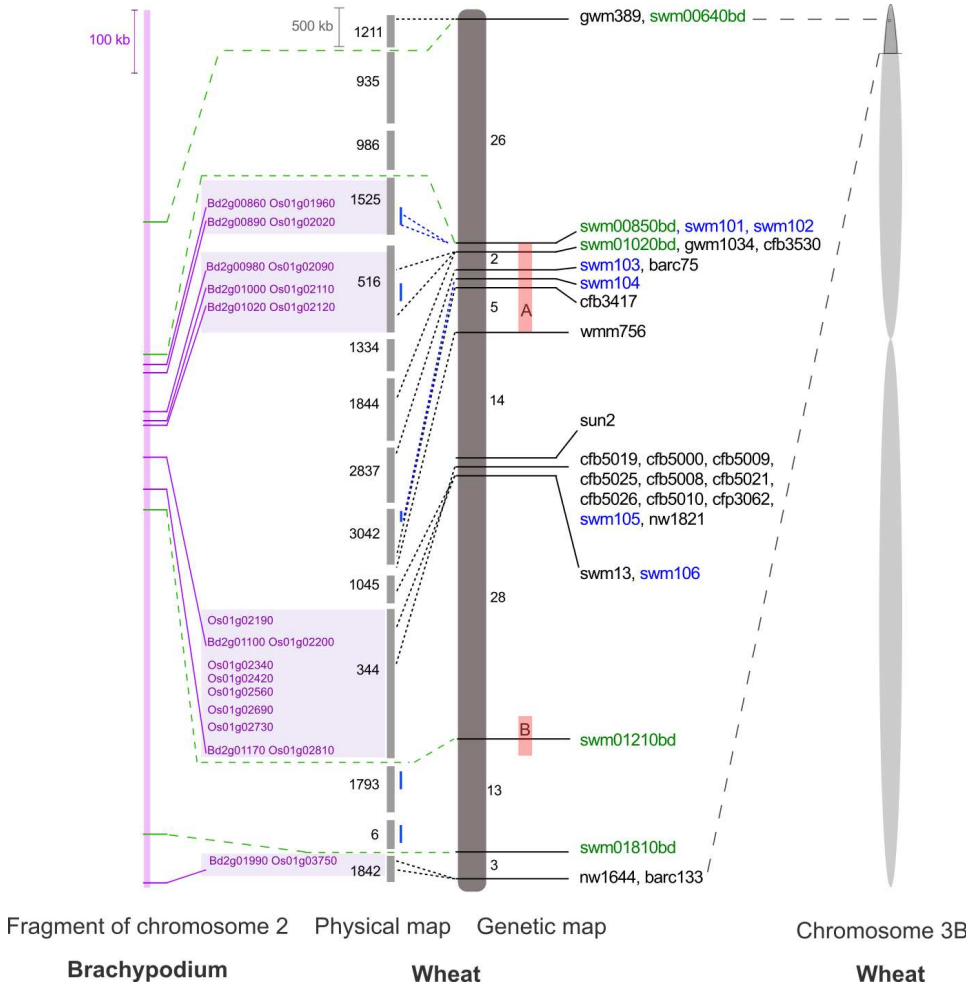

Figure 1. Comparative analysis of the target region for *Qsng.sfr-3BS* on the genetic map of the NIL population, the physical map of wheat chromosome 3B and the corresponding syntenic region in *Brachypodium distachyon*

The genetic map of the fine-mapping NIL population (92 lines) of 'Arina' x 'Forno' (**in dark grey**) is shown: numbers represent the number of identified recombinants between given markers, **in blue** – markers developed based on the sequence analysis of BAC-clones and a partial sequence of ctg344, **in green** – markers developed from SNPs and InDels identified between sequences up- and downstream of coding regions in sequenced flow-sorted chromosomes 3B of Arina and Forno; numbers in the "green" marker names correspond to the numbers of *B. distachyon* gene homologs. Physical map of wheat cv. 'Chinese Spring': **light grey bars** represent physical contigs with their numbers on the left side, short **blue bars** represent sequenced BAC-clones from corresponding physical contigs. Syntenic relationships between the wheat physical contigs and chromosome 2 of *B. distachyon* are shown in the **violet boxes** (including homologous genes from rice and *B. distachyon*). Syntenic connections between developed gene-based markers from the genetic map and their *B. distachyon* homologs are depicted with the **dotted green lines.** The **red bars (A)** – represents the interval significantly associated with both SNG resistance and plant height, **(B)** – the interval associated with SNG resistance only. Wheat chromosome 3B is shown in **light grey** with the target interval indicated in **darker grey shade**.

To increase the density of genetic map we combined the information from the physical map of chromosome 3B version 1 (Paux *et al.* 2008) and its updated version 2 (Rustenholz *et al.* 2011) to identify synteny between the wheat physical contigs and the genomes of rice (*Oryza sativa* L.) and *B. distachyon* (The International *Brachypodium* Initiative, 2010) (Figure 1). The target interval on the physical map of wheat chromosome 3B contains regions corresponding to the interval *Os01g01960-Os01g03750* in rice and *Bd2g00860-Bd2g01990* in *B. distachyon*. However, according to the analysis of Paux *et al.* (2008) the telomeric part of the physical map (ctg986, ctg935 and ctg1211) did not reveal any syntenic homologous information and the corresponding syntenic regions in rice and *B. distachyon* might be larger in this distal part.

We sequenced five BAC-clones from the physical contigs of the target region on chromosome 3B (TaaCsp3BFhA_0112H11 from ctg3042, TaaCsp3BFhA_0078F20 from ctg1525, TaaCsp3BFhA_0027A18 from ctg516, TaaCsp3BFhA_0163M03 from ctg6 and TaaCsp3BFhA_0091M22 from ctg1793). These BAC-clones were selected based on their position on the physical map in order to cover the large gaps on the genetic map (blue lines, Figure 1). The BAC sequences were used in two approaches for developing new markers. First, we designed primers to amplify the regions containing microsatellites to develop SSR markers. Second, non-repetitive regions up- and downstream of the coding sequences were amplified from both parents to develop insertion-deletion (InDel) and cleaved amplified polymorphic sequences (CAPS) markers. Three new InDel markers (*swm101, swm102, swm103*) and one CAPS marker (*swm104*) were obtained by this approach.

To further increase marker density, we sequenced flow-sorted chromosomes 3B of both parental cultivars and this sequence information was used to design single nucleotide polymorphism (SNP) and InDel markers as described in Shatalina *et al.* (2013). We selected only contigs containing genes which corresponded to the *B. distachyon* homologs from the region *Bd2g00010-Bd2g02000* of chromosome 2, showing synteny with the target region of wheat chromosome 3B. InDel and SNP markers were based on small 2-3 bp polymorphisms in intergenic and coding regions of sequences between the two cultivars as well as on the presence and absence of gene fragments. As a result, five new markers (*swm00640bp, swm00850bd, swm01020bd, swm01210bd* and *swm01820bd*) were identified and integrated in the genetic map. After integration of all newly developed markers, a total number of 31 genetic markers was mapped to the QTL target interval. The two largest regions of the map not covered by markers corresponded to 26 and 28 recombinants in the NIL population.

To estimate the ratio between genetic and physical distances in the QTL interval, we compared the genetic map with the physical map of chromosome 3B and the syntenic region of *B. distachyon*. The interval covered 11 cM on the genetic map of the NIL 'Arina' x

'Forno' population and corresponded to the 14 physical contigs from the 3B physical map of cultivar Chinese Spring (physical map version 2, Rustenholz *et al.* 2011; http://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_pub/) (Figure 1). Four physical contigs ctg1525, ctg516, ctg344 and ctg1842 (Figure 1, violet boxes) had syntenic links to the physical interval in rice and *B. distachyon* (combined data from Paux *et al.* 2008 and 3B physical map v2). The newly developed markers *swm00640bp, swm00850bd, smw01020bd, swm01210bd* and *swm01810bd* were derived from the contigs containing coding sequences homologous to *B. distachyon* genes and, therefore, they provide connections between the genetic map of wheat and chromosome 2 of *B. distachyon*. These markers support the earlier detected syntenic relationships and indicate that the most distal starting point of the syntenic interval in *B. dystachyon* is at least at *Bd2g00640*. The largest part of the physical interval – eight contigs (ctg1525-ctg344) are flanked by the coding sequences homologous to *Bd2g00850* and *Bd2g01170,* corresponding to a relatively short region on chromosome 2 of *B. dystachyon* (207 kb). These eight contigs represent roughly one third of the genetic distance in wheat. The two large gaps on the genetic map representing 26 and 28 recombination events correspond to a maximum of four physical contigs (2167 kb) and three physical contigs (2898 kb) in wheat, respectively. However, the genetic position of the marker *swm01210bd* indicates that the syntenic region corresponding to the interval between *swm13* and *swm01210bd* was short (*Bd2g01170* – *Bd2g01210*) and spanned 31 kb on chromosome 2 of *B. distachyon*. This finding suggests the presence of a recombination hot-spot in this interval of the genetic map or an insertion in the wheat region nonsyntenic to *B. distachyon*. Overall, the orthologous region of the wheat target interval in *B. dystachyon* is at least 1 Mb long and contains 135 genes.

NIL population segregated for plant height and SNG resistance

The NIL population was tested for resistance to Stagonospora nodorum glume blotch (SNG) under field conditions during the years 2011 and 2012 at three locations in Switzerland: Reckenholz, Changins and Vouvry (Table 1). We estimated the infection quality based on good differentiation of the disease scores of the parental cultivars (Table 2). In 2011, the infection developed well on the field sites in Vouvry and Changins (average AUDPC scores of 'Forno' were 526.5 and 423 compared to 'Arina' 153.3 and 93, respectively), whereas in 2012 the Reckenholz environment supported a good infection. Three other datasets (Reckenholz 2011, Vouvry 2012 and Changins 2012) were excluded from the analysis due to the insufficient development of the infection. Therefore, we included in the analysis only the phenotypic data of 67 NILs from the environments Changins and Vouvry for the infection tests in 2011 and data of 28 NILs from Reckenholz in 2012 (Figure 2).

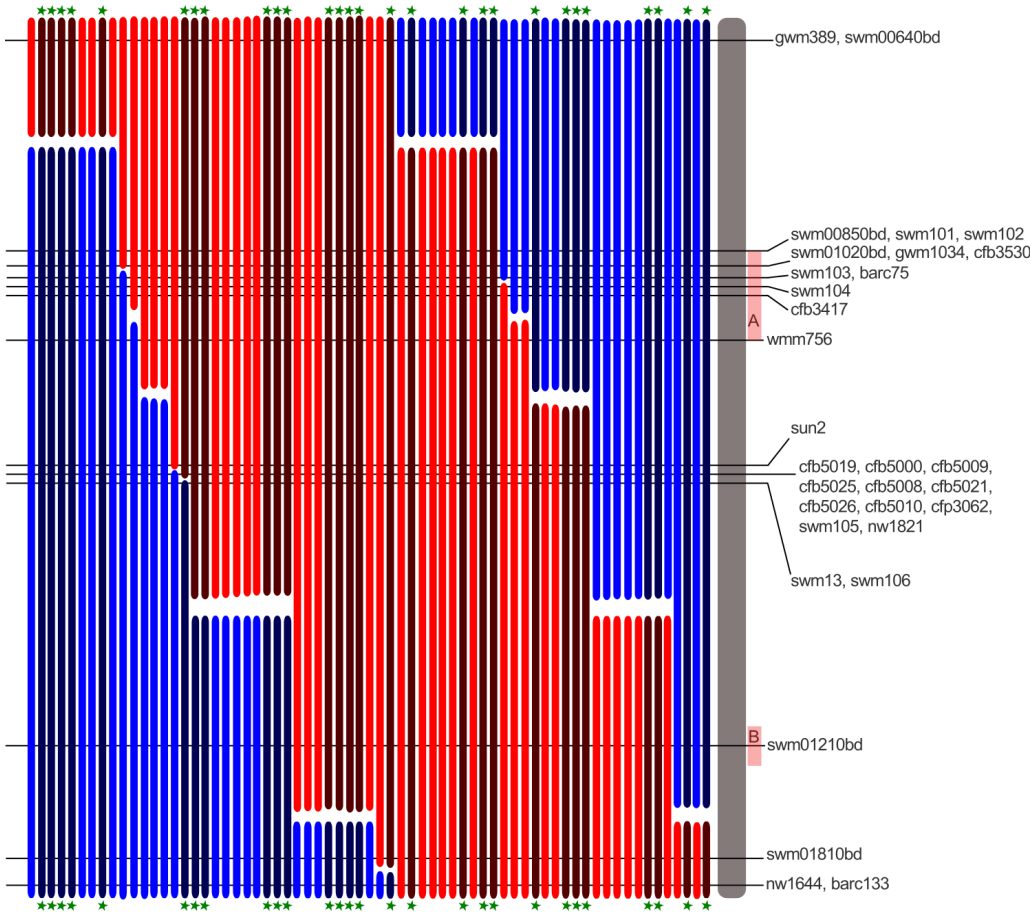

Figure 2. Graphical genotypes of the 67 and 28 NILs tested in the field

The genetic map shown on the right is based on a total of 92 NILs. **In blue** – 'Arina' allele, **in red** – 'Forno' allele. The **dark colored blue and red bars** with the **green asterisks** indicate the 28 NILs tested in field trials in 2012. **Red bars** on the genetic map are indicating the intervals associated with SNG resistance **(B)** and SNG resistance and plant height **(A)**.

Table 2. Plant height and AUDPC values of parental lines 'Arina' and 'Forno' and fine-mapping NIL population

Trait	Environment	No of NILs tested	QTL donor line Arina	Recurrent parent Forno	Average NIL	Min NIL	Max NIL
AUDPC	Vouvry-2011	67	153.3±30.8	526.5±76.5	460.4±78.5	295	701.5
	Changins-2011	67	93±8.7	423±73.6	329.4±76.9	187	575
	Reckenholz-2012	28	67.4±6.4	224.3±68.7	143.2±60.1	59	295.5
Plant height (cm)	Reckenholz-2011	67	115±7.06	103.3±5.6	103.2±5.8	90	118
	Changins-2011	67	109.9±2.4	101.8±4.4	101.6±4.6	90	115
	Reckenholz-2012	28	124.4±1.9	107.6±1.4	109.4±4.9	100	120

The last three columns show average, minimum and maximum values of the corresponding trait in the NIL population.

To test whether the SNG resistance phenotypes were reproducible under different field conditions, we calculated the Spearman correlation coefficient (r) between the replications of the same genotypes within the individual field site as well as correlation between the different fields (Table 3). The datasets from the three environments with good disease development were used: the dataset of 67 NILs (phenotypic data from 2011 from two sites), and the subset of 28 NILs with data from the Reckenholz site 2012. The correlation of the AUDPC values between the NIL replications within individual environments varied from 0.47 (Changins-2011) to 0.81 (Reckenholz-2012) and between the environments from 0.56 (correlation between Changins-2011 and Vouvry-2011) to 0.84 – correlation between Reckenholz-2012 and Changins-2011. Moderate to high correlation of the SNG resistance scores indicated their reproducibility across the environments. Due to the continuous distribution of AUDPC values, the NILs could not be grouped in two contrasting disease response classes (resistant or susceptible). This observation suggested that the disease response in the NIL population has to be analyzed as a quantitative trait and not as a single Mendelian gene.

In addition to a segregating disease response, we observed a segregation for plant height in our high-resolution mapping population. While several lines were taller than the recurrent parent of the NILs (cultivar 'Forno'), and more similar to SNG-QTL donor cultivar 'Arina', some NILs were even shorter than cultivar 'Forno'. Overall, the tested lines showed a distribution of plant height from 90 cm to 120 cm (Table 2). The rank correlation (r) of plant height and resistance to SNG ranged from -0.53 (average AUDPC and average plant height in Changins 2011) to -0.78 (average AUDPC and plant height in Reckenholz 2012) and revealed a negative correlation between these two traits (Table 3). This trend is depicted on the plot of SNG resistance (AUDPC value) against plant height of 67 NILs (Figure 3A; $R^2 = 0.20$), and becomes more pronounced in the combined dataset of 28 NILs

from Reckenholz-2012, Changins-2011 and Vouvry-2011 (Figure 3B; $R^2 = 0.59$). The rank correlation of plant height between the environments varied from a moderate correlation of 0.36 in 67 NILs to a high correlation of 0.77 in 28 NILs (Table 3).

Broad-sense heritability of SNG resistance was 0.71 for 67 NILs tested in 2011 in two environments – Changins and Vouvry. Heritability of plant height was 0.63 for 67 NILs, tested in Changins-2011 and Reckenholz-2011.

Table 3. Spearman correlation coefficients between the average values of AUDPC and plant height (PH) from different environments

No of lines			**correlation**	***p*-value**
67 NILs	AUDPC Ch-2011	AUDPC Vo-2011	0.56	<0.0001***
	PH Ch-2011	PH Re-2011	0.36	0.0016**
	AUDPC Ch-2011	PH Re-2011	-0.32	0.0047**
	AUDPC Ch-2011	PH Ch-2011	-0.53	<0.0001***
	AUDPC Vo-2011	PH Re-2011	-0.44	<0.0001***
	AUDPC Vo-2011	PH Ch-2011	-0.47	<0.0001***
28 NILs	AUDPC Ch-2011	AUDPC Vo-2011	0.77	<0.0001***
	PH Ch-2011	PH Re-2011	0.64	0.0002***
	AUDPC Ch-2011	PH Re-2011	-0.63	0.0004***
	AUDPC Ch-2011	PH Ch-2011	-0.71	<0.0001***
	AUDPC Vo-2011	PH Re-2011	-0.62	0.0004***
	AUDPC Vo-2011	PH Ch-2011	-0.66	0.0001***
	AUDPC Re-2012	PH Re-2012	-0.71	<0.0001***
	AUDPC Re-2012	PH Ch-2011	-0.81	<0.0001***
	AUDPC Re-2012	PH Re-2011	-0.7	<0.0001***
	AUDPC Re-2012	AUDPC Vo-2011	0.76	<0.0001***
	AUDPC Re-2012	AUDPC Ch-2011	0.84	<0.0001***
	PH Re-2012	PH Ch-2011	0.77	<0.0001***
	PH Re-2012	PH Re-2011	0.74	<0.0001***
	PH Re-2012	AUDPC Vo-2011	-0.54	0.0032**
	PH Re-2012	AUDPC Ch-2011	-0.59	0.001**

Significance thresholds *p<0.05, **p<0.01, ***p<0.001; Environments: Vo – Vouvry, Ch – Changins, Re – Reckenholz

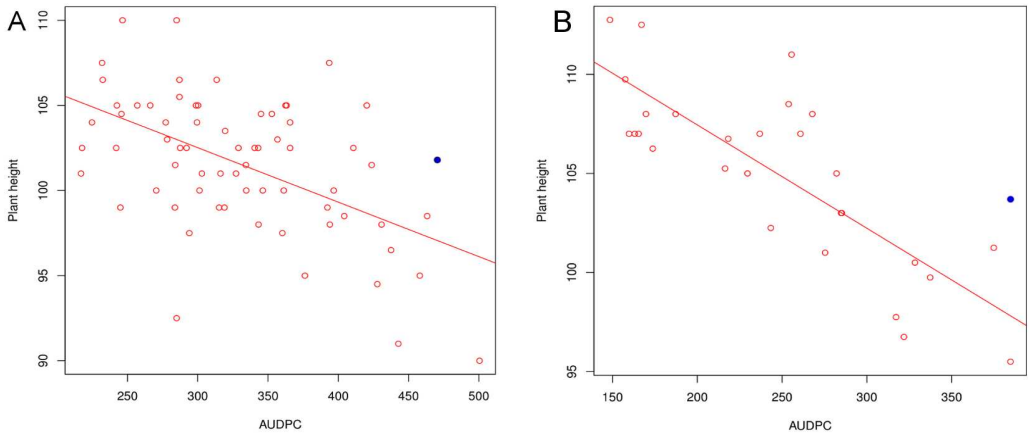

Figure 3. Average values of plant height and AUDPC score of 67 NILs (A) and 28 NILs (B)

Average values were calculated from one environment for 67 NILs and two environments for 28 NILs. Blue dots represent average value of recurrent parent (cultivar Forno).

Single marker analysis (SMA) defines regions associated with SNG resistance and plant height

To localize *Qsng.sfr-3BS* on the genetic map, we applied single marker analysis (using mixed linear model) to the two datasets (67 NILs and the subset of 28 NILs, Table 4). We included the phenotypic data from Changins-2011 and Vouvry-2011 for the analysis of 67 NILs. The SMA of this dataset detected no significant association between the AUDPC value for SNG and any of genetic markers from the target region. Analysis of the subset of 28 NILs was done using the data from Changins-2011, Vouvry-2011 and Reckenholz-2012. It identified an association between AUDPC and the interval *swm00850bd – wmm756* (region A, Figure 1). According to the physical map of chromosome 3B (Figure 1), region A has a size of 4.5 Mb (6 physical contigs).

Since the NIL population segregated for plant height, we checked whether plant height is associated with any of the markers in the target region of *QSng.sfr-3BS*. Similarly to the SMA of AUDPC, we analyzed two datasets of 67 and 28 NILs. For the analysis of 67 NILs we included the data from Changins-2011 and Reckenholz-2011. Analysis of mixed linear model detected a significant association between the trait and the genetic interval *swm00850bd – wmm756* (region A, Figure 1), with the most significant F-value for the

marker *wmm756* (Table 4). The analysis of 28 NILs included the data from Reckenholz-2012 and 2011, Changins-2011 and revealed associations of plant height with the genetic interval *gwm389* – *wmm756*, as well as with the marker *nw1644* which co-segregates with the proximal flanking marker *barc133*. We consider the association of *barc133* to be an artifact, because the mapping population consisted only of lines recombining in the target region. Therefore, the parental alleles for *barc133* were exactly opposite to the parental alleles of *gwm389* in all recombinants. Thus, as there is a lower AUDPC score of the 'Arina' allele for marker *gwm389*, there has also to be a lower AUDP value for the 'Forno' allele at marker *barc133*.

Since plant height correlated with disease resistance, the next step was to conduct SMA of disease resistance using plant height as a cofactor in the analysis. Two datasets were analysed: 67 NILs with the phenotypic data from Changins-2011 and 28 NILs with the data from Changins-2011 and Reckenholz-2012, where both AUDPC and plant height were measured (Table 5). Surprisingly, SMA for 67 NILs detected one significant association of SNG resistance with the marker *swm01210bd*, which was not detected before in SMA of AUDPC without plant height as a cofactor. Thus, SMA indicates the presence of two intervals associated with resistance to SNG within the target region for *QSng.sfr-3BS* (regions A and B, Figure 1).

The analysis of 28 NILs detected no significant association between disease resistance and any of the markers in the QTL target region, however, the genetic interval *swm00850bd* – *nw1821* had *t*-values close to significance. For both datasets (28 and 67 NILs) plant height as a fixed cofactor in the model was highly significant. The calculated percentage of the total variance of SNG AUDPC explained by plant height was 34.5% for 67 NILs. Therefore, plant height and disease resistance to SNG for the region A might have the same genetic basis.

Table 4. Single marker analysis of SNG resistance and plant height

Marker	Plant height				AUDPC			
	67 NILs[a]		28 NILs[b]		67 NILs[c]		28 NILs[d]	
	F-value[e]	*p*-value[f]	*F*-value[e]	*p*-value[f]	*F*-value[e]	*p*-value[f]	*F*-value[e]	*p*-value[f]
gwm389	1.324	0.2508	5.593	0.0192*	0.369	0.544	1.250	0.2653
swm00640bd	1.324	0.2508	5.593	0.0192*	0.369	0.544	1.250	0.2653
swm00850bd	5.233	0.023*	6.172	0.014*	0.359	0.5498	6.498	0.0117*
swm101	5.233	0.023*	6.172	0.014*	0.359	0.5498	6.498	0.0117*
swm102	5.233	0.023*	6.172	0.014*	0.359	0.5498	6.498	0.0117*
swm01020bd	5.233	0.023*	6.172	0.014*	0.359	0.5498	6.498	0.0117*
gwm1034	5.233	0.023*	6.172	0.014*	0.359	0.5498	6.498	0.0117*
cfb3530	5.116	0.0245*	6.172	0.014*	0.637	0.4256	6.498	0.0117*
swm103	5.598	0.0187*	6.172	0.014*	0.177	0.6742	6.498	0.0117*
barc75	5.598	0.0187*	6.172	0.014*	0.177	0.6742	6.498	0.0117*
swm104	6.239	0.0131*	6.172	0.014*	0.004	0.9501	6.498	0.0117*
cfb3417	6.366	0.0122*	6.172	0.014*	0.069	0.7935	6.498	0.0117*
wmm756	9.126	0.0028**	6.172	0.014*	1.171	0.2803	6.498	0.0117*
sun2	2.196	0.1396	0.131	0.7178	0.025	0.8741	1.086	0.2989
cfb5000	2.814	0.0947	0.131	0.7178	0.615	0.4336	1.086	0.2989
cfb5009	2.814	0.0947	0.131	0.7178	0.615	0.4336	1.086	0.2989
cfb5025	2.814	0.0947	0.131	0.7178	0.615	0.4336	1.086	0.2989
cfb5008	2.047	0.1538	0.131	0.7178	0.123	0.7262	1.086	0.2989
cfb5021	2.814	0.0947	0.131	0.7178	0.615	0.4336	1.086	0.2989
cfb5026	2.814	0.0947	0.131	0.7178	0.615	0.4336	1.086	0.2989
cfb5010	2.814	0.0947	0.131	0.7178	0.615	0.4336	1.086	0.2989
cfp3062	1.984	0.1601	0.131	0.7178	0.600	0.4395	1.086	0.2989
nw1821	2.814	0.0947	0.131	0.7178	0.615	0.4336	1.086	0.2989
swm106	3.976	0.0472*	1.229	0.2693	0.353	0.5528	0.966	0.3271
swm01210bd	2.248	0.135	2.825	0.0948	1.177	0.279	0.058	0.8100
swm01810bd	0.24	0.6246	2.384	0.1246	1.128	0.2892	0.345	0.558
nw1644	1.324	0.2508	5.593	0.0192*	0.369	0.544	1.250	0.2653
barc133	1.324	0.2508	5.593	0.0192*	0.369	0.544	1.250	0.2653

a – SMA (mixed linear model) of the 67 NILs with two replications from 2 environments (Changins-2011 and Reckenholz-2011)
b – SMA (mixed linear model) of the 28 NILs with two replications from 3 environments (Changins-2011, Reckenholz-2011 and Reckenholz-2012)
c – SMA (mixed linear model) of the 67 NILs with two replications from 2 environments (Changins-2011 and Vouvry-2011)
d – SMA (mixed linear model) of the 28 NILs with two replications from 3 environments (Changins-2011, Vouvry-2011 and Reckenholz-2012)
e – *F*-value of the given marker effect
f – Significance thresholds of the given marker effect *p<0.05, **p<0.01, ***p<0.001

Table 5. Single marker analysis of SNG resistance with plant height as a cofactor

Marker	67 NIL[a]		28 NIL[b]	
	t-value[c]	*p*-value[d]	*t*-value[c]	*p*-value[d]
gwm389	-0.519	0.605	0.886	0.378
swm00640bd	-0.519	0.605	0.886	0.378
swm00850bd	0.007	0.995	1.960	0.0526
swm101	0.007	0.995	1.960	0.0526
swm01020bd	0.007	0.995	1.960	0.0526
gwm1034	0.007	0.995	1.960	0.0526
cfb3530	0.020	0.984	1.960	0.0526
swm103	0.180	0.858	1.960	0.0526
barc75	0.180	0.858	1.960	0.0526
swm104	0.431	0.667	1.960	0.0526
cfb3417	0.168	0.867	1.960	0.0526
wmm756	-0.119	0.905	1.960	0.0526
sun2	-0.171	0.865	1.978	0.0504
cfb5000	0.273	0.786	1.978	0.0504
cfb5009	0.273	0.786	1.978	0.0504
cfb5025	0.273	0.786	1.978	0.0504
cfb5008	0.328	0.743	1.978	0.0504
cfb5021	0.273	0.786	1.978	0.0504
cfb5026	0.273	0.786	1.978	0.0504
cfb5010	0.273	0.786	1.978	0.0504
cfp3062	0.419	0.676	1.978	0.0504
nw1821	0.273	0.786	1.978	0.0504
swm106	-0.183	0.855	1.160	0.249
swm01210bd	2.334	0.0212*	0.911	0.364
swm01810bd	0.074	0.941	-0.803	0.424
nw1644	0.519	0.605	-0.886	0.378
barc133	0.519	0.605	-0.886	0.378

a – SMA (mixed linear model) of the 67 NILs with two replications from Changins-2011
b – SMA (mixed linear model) of the 28 NILs with two replications from 2 environments (Changins-2011 and Reckenholz-2012)
c – *t*-value of the given marker effect
d – Significance thresholds of the given marker effect *p<0.05, **p<0.01, ***p<0.001

Two genetically distinct regions are contributing to SNG resistance in NILs

SMA indicated the presence of two loci associated with SNG resistance. We wanted to study if additional analysis could further substantiate this finding. Therefore, we analyzed the distributions of the average AUDPC and plant height of two environments for 67 NILs in the two allele classes for individual markers (Table 6). Marker *wmm756* showed the largest height difference of 2 cm for 'Arina' and 'Forno' classes, supporting the SMA described above. The analysis of the allele classes for AUDPC values revealed relevant differences for markers *wmm756, swm01210bd* and *swm01810bd*. The 'Arina' class had the lowest AUDPC values in the groups of markers *wmm756* from the region A and *swm01210bd* from the region B, which corresponded to the resistance loci identified by SMA.

Table 6. Plant height and AUDPC for 67 NILs (average of two environments) divided into 'Arina' and 'Forno' classes for each of the markers (average value of the class is shown)

Marker	AUDPC		Plant height	
	Forno	Arina	Forno	Arina
gwm389	391.9	398.0	102.0	102.8
swm103	393.0	396.6	101.7	103.2
wmm756	399.4	388.6	101.5	103.5
cfb5000	389.9	398.0	101.9	103.0
swm01210bd	398.4	387.7	102.9	101.9
swm01810bd	400.8	389.9	102.2	102.5
nw1644	398.0	391.9	102.8	102.0

Next, we divided the 67 NILs into four groups based on their genotype for the markers *wmm756* and *swm01210bd* (Figure 4). The distributions of the phenotypic values of the four groups AA – 'Arina' alleles for the two markers *wmm756* and *swm01210bd*, AF – 'Arina' for *wmm756* and 'Forno' for *swm01210bd,* FA – 'Forno' *wmm756* and 'Arina' *swm01210bd* and FF – 'Forno' are shown on the Figure 4. A *t*-test did not detect significant differences between the groups, but when considering the group means there was a clear trend. The average values from the two environments Changins-2011 and Reckenholz-2011 for plant height showed that the AA and AF groups were taller (Figure 4A). A smaller size of boxes and whiskers on the plot for the AA and AF groups in comparison to the FA and FF groups also indicated that plant height values were distributed within a smaller range in the first two groups with the 'Arina' allele of marker *wmm756*. The FA and

FF groups with the 'Forno' allele of *wmm756* showed a wider distribution of plant height which is reflected in larger boxes and longer whiskers on the plot. The larger size of the boxes for the FA and FF groups might indicate that in the absence of *wmm756* 'Arina' allele plant height was more strongly influenced by environmental factors, creating a wider distribution of phenotypic values.

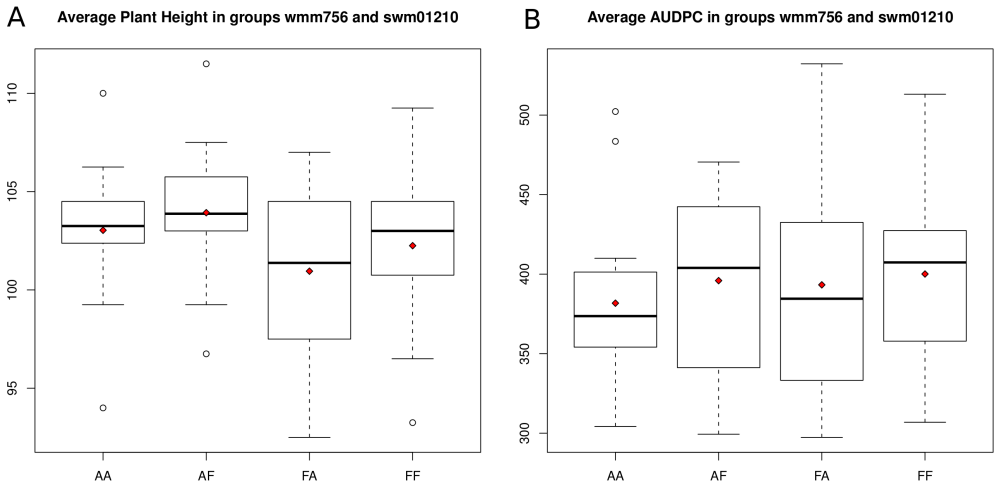

Figure 4. Box-plots of the average AUDPC and plant height of 67 NILs in four genotypic groups

67 NILs were divided in 4 groups based on genotypes of two markers (*wmm756* and *swm01210bd*). **AA** – 'Arina' allele for *wmm756* and 'Arina' allele for *swm01210bd*, **AF** – 'Arina' *wmm756* and 'Forno' *swm01210bd*, **FA** – 'Forno' *wmm756* and 'Arina' *swm01210bd* and **FF** – both markers have 'Forno' allele. The **black line** is the median of each group and the **red diamond-shaped dot** represent the group mean. **(A)** average plant height among 2 environments, **(B)** average AUDPC score. The whiskers depict the minimum and maximum values.

In agreement with the hypothesis of two resistance loci present in the target region, the average AUDPC values of 67 NILs from the two environments Changins-2011 and Vouvry-2011 were lowest for the AA group (Figure 4B). The short length of the upper whisker for the AA group suggested that except for the two outliers, this group showed a clear tendency for a stronger resistance as the AUDPC values fall into the same range of low values. In contrast, the FF group showed the highest AUDPC of both median and mean values among all genotypic groups. The sizes of boxes and whiskers for the AF, FA and FF groups were larger than for the AA group, which indicated a wider distribution of the AUDPC values in these groups. The wider distribution of the AUDPC in the groups with no (FF) or only one resistance locus (AF and FA) suggested that the presence of two resistance loci (group AA) indeed resulted in strongest resistance. The AF, FA and FF

groups might be more influenced by other genetic components and environmental factors. Thus, we conclude that there were two resistance loci in the target interval and the presence of both loci provided the strongest resistance response. The region A resistance locus was associated with both plant height and SNG resistance. The effect of the resistance locus B linked to *swm01210bd* seemed to be larger than the effect of the resistance locus A linked to *wmm756*.

Discussion

Field infection tests of a NIL population revealed a segregation in quantitative manner for both SNG resistance and plant height. The two traits negatively correlated with moderate to high levels and they were reproducible between different tested environments. Further analysis of the target region identified two loci: region A (*swm00850bd – wmm756*) and B (*swm01210bd),* associated with SNG resistance. Region A was also associated with plant height.

Construction of a genetic map in the 'Arina' x 'Forno' NIL population

The physical map of chromosome 3B allowed us to improve the genetic map for the target interval of *QSng.sfr-3BS* by integrating recently developed markers as well as by establishing additional markers. In contrast to the results in other populations, we detected a very low level of polymorphism in ISBP markers in our NIL population (Paux *et al.* 2010). In fact, ISBP markers are based on differences in the insertion sites of (retro)transposons and since our NIL population is derived from a cross of two Swiss winter varieties, there might be less diversity in these repeat junctions than in 'Chinese Spring' x 'Renan', 'Courtot' x 'Chinese Spring' or ITMI populations, which were used for mapping of ISBPs to anchor the physical map of chromosome 3B. Another problem was the presence of gaps in the genetic map (regions not covered with genetic markers), which could not be predicted or explained by the physical map or synteny with *B. distachyon.* Those gaps might be recombination hot spots (Faris *et al.* 2000; Akhunov *et al.* 2003) or large chromosomal insertions specific to the 'Arina' x 'Forno' population. A larger number of mapped markers will be needed to further test the “hot spot”- hypothesis. Thus, recently developed genomic resources can be helpful for high resolution mapping purposes, but individual differences between the parental lines and the cultivars used for the construction of the physical map have to be considered. Sequenced flow-sorted

chromosomes 3B of the parental lines for the NIL population provided an important source of information to increase the marker density. SNP markers derived from sequences of the particular parental cultivars combined with the synteny information are very specific and enabled us to target selected regions. Synteny with *B. distachyon* allowed us to predict the possible position and to choose the potential markers specifically for the critical parts of the genetic map. Both resources, the physical map and the sequence information from 'Arina' and 'Forno', might be extremely valuable for identification of candidate genes when the target interval is further narrowed.

SNG resistance is strongly influenced by environmental conditions

Successful development of SNG infection is dependent on environmental conditions such as temperature and humidity (Eyal *et al.* 1987; De Wolf and Francl 2000; Shah and Bergstrom 2000). In agreement with these earlier observations, among 6 conducted field infection tests (2 years, 3 environments) only three trials had optimal conditions for disease development and three datasets were, therefore, discarded from the analysis. Environmental dependence was further suggested by the analysis of variance, which showed the significant impact of environmental factors (58.6% of the variance) on the disease development. The absence of a significant SMA result for AUDPC values in the dataset of 67 NIL further indicates a strong environmental influence on the trait. Therefore, the range of correlation coefficients of AUDPC values (0.47 to 0.81) between and within individual environments indicated on the one hand, presence of the proposed QTL in the selected NILs, and on the other hand, a strong influence of environmental factors on SNG resistance. Reproducibility of the results were demonstrated by the correlation between three environments (0.56 to 0.84) and heritability of the trait (0.71 for 67 NILs). These results are very similar to the results observed previously in the 'Arina' x 'Forno' RIL population: correlation of SNG resistance between environments varied from 0.59 to 0.84 and heritability of the trait was 0.80 (Schnurbusch *et al.* 2003). These results from phenotyping show that our data are reliable, but to variable for clearly assigning a genotype to a “resistant” or “susceptible” class. Future experiments to obtain resistance data should possibly be done at a larger number of field sites and over several years to get data of sufficient quality for high-resolution mapping and ultimately for cloning of the resistance gene(s).

Is the SNG resistance QTL linked to *wmm756* a plant height QTL?

In addition to the segregation for SNG resistance, we observed segregation for plant height in our NIL population. This segregation was reproducible and heritable, and SNG resistance and plant height showed a moderate to high negative correlation in all tested

environments. A correlation between plant height and resistance to SNG was previously shown under field conditions suggesting disease escape for the tall plants (Broennimann *et al.* 1973; Scott *et al.* 1982; Van Ginkel and Rajaram 1999; Wicki *et al.* 1999). Association of plant height and SNG resistance was also identified in the previous QTL study on the RIL population of 'Arina' x 'Forno' (Schnurbush *et al.* 2013) and the RIL population of 'Forno' x 'Oberkulmer' (Aguilar *et al.* 2005). However, the QTLs for plant height detected by Schnurbusch *et al.* (2003), were not located in the target region for *QSng.sfr-3BS*, but on chromosomes 5BL, 5AL, 7BS, 3BL and 2AL on the genetic map of 'Arina' x 'Forno' RILs . The *QSng.sfr-3BS* interval did not show any significant association with other traits such as plant height or flowering time (Schnurbusch *et al.* 2003). The fact that we detected a region associated with plant height in the NIL population which was not detected in RILs might be explained by epistatic effects of other plant height QTLs in the RIL 'Arina' x 'Forno' population.

Segregation for plant height and its correlation with SNG resistance in our NIL population was unexpected: we used artificial infection after spike emergence in the field and did not rely on natural infections. Artificial infection by spraying plants from above should have decreased the disease escape due to plant height, because in these conditions tall and short plants received equal amount of spores directly to the spikes in contrast to natural infection which usually is spreading from infected ground debris from the lower leaves to the ears by rain splashes. Nevertheless, tall and short plants could also differ in microclimate conditions of the spikes which are known to be important for disease development (Scott *et al.* 1985). Measuring of additional phenotypic traits such as spike morphology or distance between spikes and flag leaves in future field trials might help to clarify whether plant height *per se* is influencing SNG resistance or if any other minor features of the tested NILs contribute to the disease escape.

Further evidence for the correlation of plant height and SNG resistance was found by including plant height as a factor in general linear model analysis which significantly influenced the SMA results. Furthermore, the previously detected significance for the genetic markers from the interval *swm00850bd – wmm756* (region A, Figure 1) became less pronounced, indicating that a large part of the variance was explained by plant height. The correlation of plant height and SNG resistance could have several explanations. SNG resistance in region A could be a gene controlling plant height or it might contain both a resistance gene as well as a plant height gene within a gene cluster, which would explain a co-segregation of the traits. However, the most likely explanation is that the region A controls plant height and results in a change of microclimate at the spike level, with SNG resistance improvement as a pleiotropic effect of the plant height gene.

Two genetically distinct SNG resistance loci are present in the *QSng.sfr-3BS* target interval

The absence of any influence from the genetic background in a NIL population theoretically allows to “Mendelize” the trait of interest and to divide the phenotypic response of the fine-mapping population in two contrasting classes – “resistant” and “susceptible” (Keurentjes *et al.* 2007; Pea *et al.* 2009). *QSng.sfr-3B* did not manifest itself as Mendelian trait in our NIL population and was scored quantitatively. The identification of two resistance loci in the *QSng.sfr-3BS* target interval as well as the strong environmental influence possibly explain the observed quantitative response to *S. nodorum* infection. The two loci were detected by SMA: both markers *wmm756* (region A) and *swm01210bd* (region B) were significantly associated with disease resistance. The analysis of genotypic groups, taking into account the allelic composition of both markers for region A and B, indicated the same tendency. However, no statistically significant differences were found between the groups in this type of analysis. This absence of significant differences could be due to the relatively low number of NILs in the four genotypic groups as well as the high environmental dependence of the resistance trait. Since the NIL population was developed based on the hypothesis of one QTL localized in the target interval, the number of recombinants was clearly not sufficient for the unexpected need analysis of two genetically distinct loci. To separate the effects of two loci and isolate each of them in the ‘Forno’ genetic background, the number of NILs has to be at least doubled in comparison to the 67 lines analyzed in our study. Additionally, some variation among the NILs for both plant height and AUDPC which probably was caused by the remaining 6.25% genetic background from 'Arina' might have also influenced the results of the analysis.

Tightly linked QTLs controlling the same trait have previously been described for other quantitative traits. In tomato, two QTLs *Brix9-2-5* and *PW9-2-5* located 0.3 cM apart are both modifying sugar content (Fridman *et al.* 2002). In rice, QTL *Hd3* consisted of two linked QTLs *Hd3a* and *Hd3b* controlling heading date (Monna *et al.* 2002). A QTL for flowering time *dth1.1* consisted of at least two distinct sub-QTLs *dth1.1a* and *dth1.1b* (Thomson *et al.* 2006). Another example is a growth-rate QTL in *Arabidopsis thaliana*, consisting of two tightly linked QTLs located within a 210 kb interval (Kroymann and Mitchell-Olds 2005). The SNG resistance analysis presented in this study focusing on *QSng.sfr-3BS* revealed an additional case for two closely linked QTL. Furthermore, the pleiotropic effect of one the two involved loci on plant height further complicated the analysis. These findings indicate an unexpectedly high genetic complexity of SNG resistance in cultivar ‘Arina’ and the need for the development of larger populations where the effects of the two linked QTL are separated from each other in a sufficiently large number of individual genotypes. Only in a such a mendelized form in even more identical backgrounds a future molecular cloning can be envisaged.

Acknowledgements

We thank Bea Senger and Stefan Kellenberger for excellent technical assistance with the field experiments. We thank Philip Streckeisen for the guidance on SNB field scorings. We also thank Bea Senger for the technical help with the plant material in the greenhouse. We thank Hana Simkova and Jaroslav Dolezel for isolation of chromosome 3B by flow cytometry. We are grateful to Matthias Helling for the help with the statistical analysis. The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the grant agreement FP7-212019 (Triticeae Genome) and was supported in the framework of the European Cooperation in Science and Technology FA0604 (Tritigen). This work was also supported by a grant from the Swiss National Science Foundation 31003A_127061.

CHAPTER 4

General discussion

In this thesis we developed a new protocol to generate gene-based SNP markers using flow sorted chromosomes (Chapter 2). These markers were used to fine-map a QTL for resistance to Stagonospora nodorum glume blotch on chromosome 3B. High resolution mapping of this QTL produced a rather unexpected, but very relevant result. Interestingly, we found two genetically distinct loci (A and B) associated with SNG resistance within the small target interval (Chapter 3). One of the resistance loci, locus A, was also linked to increased plant height. This finding indicates a complex mechanism of disease resistance conferred by the target QTL.

Possible mechanisms of SNG resistance on chromosome 3B

Association of plant height with SNG resistance was detected multiple times (Broennimann *et al.* 1973; Scott *et al.* 1982; Van Ginkel and Rajaram 1999; Wicki *et al.* 1999; Schurbusch *et al.* 2003). Generally, tall plants are more resistant than shorter plants. However, since previous QTL mapping analysis in the 'Arina' x 'Forno' RIL population did not detect a QTL for plant height in the target region (Schnurbusch *et al.* 2003), it was unexpected to observe a segregation for this trait in the NIL mapping population. The analysis of phenotypic data indicated that SNG resistance and plant height were associated with the same locus A, which suggests that both traits were probably controlled by the same gene(s). Therefore, locus A might confer SNG resistance by controlling genes altering plant height. This hypothesis needs further verification with additional field phenotypic data from several environments and larger number of lines from the mapping population. Since SNG resistance in our NIL mapping population showed high environmental dependence, additional data from several trials with a sufficient progress of

S. nodorum infection could improve the power of genetic mapping and allow to determine whether SNG resistance and plant height are controlled by the same gene or by several closely linked genes at locus A. Additionally, the difference in disease progress might be caused by altered microclimate that favours development of the disease in short plants (Scott *et al.* 1985). It would be interesting to phenotype other morphological traits which might influence the spike microclimate such as a peduncle length, spike density and wax content of the spike surface. It has been shown that wheat leaf waxes influence the ability of *S. nodorum* spores to attach to a hydrophobic surface during the first step of the infection (Zellinger *et al.* 2006; Newey *et al.* 2007; reviwed in Lazniewska *et al.* 2012). Phenotyping of these traits might provide further insight into the mechanism of SNG resistance conferred by resistance locus A.

The second locus B was not linked to increased plant height and might be based on a different mechanism. It might provide SNG resistance which could be functionally similar to the toxin insensitivity shown for SNL resistance (Friesen *et al.* 2007). Since the S. nodorum leaf blotch is caused by toxin-gene interactions, phenotyping tests for the disease resistance are based on the leaf infiltrations with toxins of *S. nodorum* (Friesen *et al.* 2006; Friesen *et al.* 2007; Abeysekara *et al.* 2009). It might be useful to test the same approach for S. nodorum glume blotch on the NIL genotypes for the height-independent resistance locus B. The genetic separation of each locus from the target region would require a larger number of the NILs recombining in the interval between region A and B. A sufficient subset of NILs containing only one of the identified resistance loci might allow us to test whether resistance locus A controls plant height and influences resistance phenotype by altered microclimate of spikes and whether the locus B confers resistance based on toxin insensitivity.

Genes controlling plant height

Our analysis showed that the identified resistance locus A might control plant height. Hence, information about the genes controlling plant height might help to identify possible candidate genes in the target region. Genes controlling plant height are often related to the pathways of plant growth hormones. To date, the two major pathways of the plant growth hormones gibberellins (GA) and brassinosteroids (BR) that influence plant height are well studied in *Arabodopsis thaliana* and rice. It has been shown that both GA metabolism and signal transduction are important for control of plant height: mutants in GA signalling or synthesis result in dwarfing phenotype (reviewed in Fernandez *et al.* 2009). Semi-dwarf phenotypes in rice and wheat are a hallmark of the “green revolution”. Genes conrtrolling semi-dwarf phenotypes in wheat were introduced by N. Borlaug at the International Maize and Wheat Improvement Center (CIMMYT). Semi-dwarf genes prevent lodging of the plants in high wind and rains. They are controlled by mutations in

GA-related genes. Two of the semi-dwarfing “green revolution” genes in wheat, *Rht-B1b* and *Rht-D1b,* encode truncated DELLA proteins (Peng *et al.* 1999). DELLA proteins play an important role in GA signal transduction as repressors of GA-induced gene transcription in the absence of GA signalling (Zentella *et al.* 2007). In rice, the semi-dwarf phenotype of the “green revolution” is controlled by a mutated *semidwarf1* (*sd1*) gene which encodes a GA20 oxidase – a member of the GA synthesis pathway (Sasaki *et al.* 2002).

BRs promote plant growth because they promote cell elongation (Clouse *et al.* 1996; Li and Jin 2007; Gendron and Wang 2007; Krishina 2003; Fujioka and Yokota 2003; Bishop, 2003). In rice, characterization of BR-deficient mutants, OsDwarf2 and OsDwarf11, led to identification of novel enzymes involved in BR-biosynthesis (Hong *et al.* 2005; Tanabe *et al.* 2005). Additionally, there are genes regulating plant height which are not part of GA or BR pathways. For instance, the rice gene *SPS* encodes a sucrose phosphate synthase and controls a plant height QTL (Ishimaru *et al.* 2004). In sorghum, *Dw3,* influencing plant height, encodes a P-glycoprotein that regulates polar auxin transport (Multani *et al.* 2003).

Thus, genes involved in GA- and BR synthesis or signalling might be interesting candidate genes for the SNG resistance locus A on chromosome 3B.

Approaches to improve the genetic map of the target interval for SNG resistance QTL

The genetic map of the target region (described in Chapter 3) could be further improved by using the information from the sequences of flow-sorted chromosomes 3B of parental cultivars 'Arina' and 'Forno'. Additional marker design can include both SNP and InDel variations, as discussed in Chapter 2. There are several possible options for further marker development: (1) not all potential SNP/InDels were tested in our work, (2) the SNP coverage threshold could be lowered to test if the SNPs with less sequence coverage could also be used for marker development and (3) since we tested only SNPs located upstream and downstream of the genes (Chapter 2), SNPs and InDels within coding sequences of 'Arina'/'Forno' contigs can be tested for polymorphism in NIL population. The physical map of chromosome 3B and annotation of its sequence assembly are continuously progressing. Therefore, alignments of the annotated chromosome 3B scaffolds of cultivar 'Chinese Spring' and sequenced contigs of 'Arina'/'Forno' from the chromosomes 3B can be produced. The alignments could be done at least for the parts of 'Chinese Spring' scaffolds which contain annotated genes and 'Arina'/'Forno' contigs with identified coding sequences. This comparative sequence information might be beneficial for the marker design as well as for the future candidate gene identification.

Additionally, it might be helpful to genotype the NIL population using a new large scale 90K SNP array for wheat (described below). The output from this analysis might possibly

provide new markers for the target region as well as detailed genotypic information about the genetic background of the NILs. Since there is still 6.25% of genetic background from the donor parent 'Arina', it might be helpful to identify which chromosomal segments are from 'Arina'. This information might help to detect whether any other genomic regions, besides the target interval with the identified two resistance loci, are influencing the observed phenotypes.

High-throughput genotyping in wheat

Resequencing for mapping has been extensively used in different plant species with sequenced genomes such as *Arabidopsis thaliana* (Clark *et al.* 2007; Ossowski *et al.* 2008; Cao *et al.* 2011), rice (Huang *et al*. 2009; Xu *et al.* 2012; reviewed in Huang *et al.* 2013) and maize (Lai *et al.* 2012). For wheat, resequencing of the whole genome remains expensive. Additionally, given the large number of repetitive elements in hexaploid genome, it is not very informative for the polymorphism detection due to problems with repeat assembly.

Resequencing of part of the wheat genome might be an alternative strategy. For example, exon capture technology was proven to be efficient for large-scale discovery of polymorphisms in the human genome. The exon capture method is based on sequence capture by long oligonucleotide baits, which is performed using solid or liquid-phase hybridization assays (Albert *et al.* 2007; Porreca *et al.* 2007; Okou *et al.* 2007; Gnirke *et al.* 2009; Teer and Mullikin 2010). In maize, Fu *et al.* (2010) conducted array-based sequence capture and identified 2500 high-quality SNPs between the reference genome of line B73 and another maize line Mo17. Exon capture was also tested in soybean (Haun *et al.* 2011) and for *de novo* assemblies of the coding sequences and polymorphism detection in cotton (Salmon *et al.* 2012) and tetraploid wheat (Saintenac *et al.* 2011). Thus, sequencing of reduced representations of the whole genome is an efficient method for SNP discovery and mapping. It could possibly also be applied for the whole-genome SNP identification in gene mapping projects in bread wheat.

One of the important measurements of genetic variation used for mapping in bread wheat is SNP genotyping. Following the success of SNP markers in genotyping of maize and rice, large SNP arrays for wheat were recently designed. A 9K Infinium wheat SNP-chip incorporates the data from the large transcriptome projects in Australia and the USA (Aknunov, www.illumina.com; reviewed in Feuillet *et al.* 2012). For example, a novel wheat gene, *H34*, for resistance to Hessian fly (*Mayetiola destructor*) was identified using the 9K Infinium wheat SNP-chip (Li *et al.* 2013). Additionally, by including newly developed EST- and cDNA-based SNPs from the UK effort for wheat improvement (Allen et *al.* 2011) and COS-SNPs from the *TriticeaeGenome* project (www.triticeaegenome.eu) a

new 90K SNP-chip has been developed (Akhunov, personal communication; reviewed in Feuillet *et al.* 2012). These arrays contain SNP markers derived from different projects including EST sequences and genomic sequences of different wheat varieties. The SNP arrays provide an opportunity to conduct genome-wide studies in a large number of cultivars or with several mapping populations.

However, SNP mapping in large polyploid genomes is challenging because of the presence of repetitive regions and very similar homeologous sequence copies. These can lead to non-specific amplifications or detection of polymorphisms between homeologous sequence copies instead of SNP identification between two cultivars/lines. The solution of this problem could be in genotyping populations using large SNP arrays with verified homeologous-specific markers. Another option to develop homeologous-specific SNP markers is to design them from regions which are less conserved than genes and therefore, have a higher potential to be chromosome-specific. At the moment, the majority of available SNP markers are based on coding sequences and ESTs, which decreases the likelihood of these SNPs to be polymorphic in a large set of cultivars. Development of SNP markers specifically for each mapping population should increase the number of polymorphic markers. We successfully applied this approach for SNP markers development using sequences of flow-sorted chromosomes of the cultivars 'Arina' and 'Forno' which resulted in 74% of polymorphic chromosome-specific markers (as described in Chapter 2).

Individual SNP assays can be particularly useful if the parental lines of the mapping population have a very similar genetic background and therefore the mapping population is not expected to show many polymorphisms using the common SNP-chip in the particular target interval. Additionally, the success of mapping of a sufficient number of SNP markers for the high-resolution analysis of a small genetic interval largely depends on the chromosomal distribution of the SNP markers included in the common SNP-chip. In contrast, genotype-specific SNP mapping may allow to enrich the particular region of the genetic map with gene-associated SNP markers selected based on synteny with other sequenced grasses (e.g. *B. distachyon*). Genotype-specific SNP mapping may become less expensive and more efficient with the release of a wheat reference genome assembly. Alignments of the cultivar-specific sequences to the reference genome assembly will possibly allow to decrease the coverage for SNPs discovery.

REFERENCES

Abeysekara NS, Friesen TL, Keller B and Faris JD (2009) Identification and characterization of a novel host-toxin interaction in the wheat-Stagonospora nodorum pathosystem. *Theor Appl Genet* **120**:117-126

Agarwal M, Shrivastava N and Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep* **27**:617-631

Aguilar V, Stamp P, Winzeler M, Winzeler H, Schachermayr G, Keller B, Zanetti S and Messmer M M (2005) Inheritance of field resistance to Stagonospora nodorum leaf and glume blotch and correlations with other morphological traits in hexaploid wheat (*Triticum aestivum* L.). *Theor Appl Genet* **111**:325-336

Akhunov E, Nicolet C and Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet* **119**:507-517

Akhunov ED, Akhunova AR and Dvorák J (2005) BAC libraries of *Triticum urartu, Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. *Theor Appl Genet* **111**:1617-1622

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo H, Huo N, Lazo GR, Luo M-C, Ma YQ, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W and Dvorak J (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* **11**:702

Akhunov ED, Akhunova AR, Linkiewicz AM, Dubcovsky J, Hummel D, Lazo G, Chao S, Anderson OD, David J, Qi L, Echalier B, Gill BS, Miftahudin, Gustafson JP, Rota ML, Sorrells ME, Zhang D, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng J, Lapitan NLV, Wennerlind EJ, Nduati V, Anderson JA, Sidhu D, Gill KS, McGuire PE, Qualset CO and Dvorak J (2003) Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. *Proc Natl Acad Sci USA* **100**:10836-10841

Akhunov ED, Goodyear AW, Geng S, Qi L-L, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S, Anderson OD, Linkiewicz AM, Dubcovsky J, Rota ML, Sorrells ME, Zhang D, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng J, Lapitan NLV,

Gonzalez-Hernandez JL, Anderson JA, Choi D-W, Close TJ, Dilbirligi M, Gill KS, Walker-Simmons MK, Steber C, McGuire PE, Qualset CO and Dvorak J (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res* **13**:753-763

Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM and Gibbs RA (2007) Direct selection of human genomic loci by microarray hybridization. *Nat Methods* **4**:903-905

Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N and Edwards KJ (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat *(Triticum aestivum* L.*)*. *Plant Biotechnol J* **9**:1086-1099

Arseniuk E, Czembor PC, Czaplicki A, Song Q, Cregan PB, Hoffman DL and Ueng PP (2004) QTL controlling partial resistance to Stagonospora nodorum leaf blotch in winter cultivar Alba. *Euphytica* **137**:225-231

Balter M (2007) Plant science. Seeking agriculture's ancient roots. *Science* **316**:1830-1835

Barker GLA and Edwards KJ (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. *Plant Biotechnol J* **7**:318-325

Bennett MD and Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philos Trans R Soc Lond B Biol Sci* **274**: 227-274

Bishop GJ (2003) Brassinosteroid Mutants of Crops. *J Plant Growth Regul* **22**:325-335

Bossolini E, Wicker T, Knobel PA and Keller B (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. *Plant J* **49**:704-717

Bostwick DE, Ohm HW, Shaner G (1993) Inheritance of Septoria glume blotch resistance in wheat. *Crop Sci* **33**:439–443

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo M-C, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KFX, Edwards KJ, Bevan MW and Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**:705-710

Broennimann A (1968) Untersuchungen über *Septoria nodorum* Berk. des Weizens. *Mitteil Schweiz Landw* **16**:65–76

Broennimann A (1975) Contributions to genetics of tolerance towards *Septoria nodorum* Berk. in wheat (*Triticum aestivum* L.). *J Plant Breed* **75**:138–160

Broennimann A, Fossati A, Hani F (1973) Spreading of *Septoria nodorum* Berk. and

damage to artificially induced mutants of winter wheat cultivar Zenith (*Triticum aestivum* L.). *J Plant Breed* **70**:230–245

Bryan GJ, Stepherson P, Collins A, Kirby J, Smith JB, Gale MD (1999) Low level of DNA sequence variation among adapted genotypes of hexaploid wheat. *Theor Appl Genet* **99**:192–198

Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Töpsch S, Vos P, Salamini F and Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* **88**:695-705

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid KJ and Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* **43**:956-963

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo M-C and Dubcovsky J (2009) Analysis of gene-derived SNP marker polymorphism in US wheat *(Triticum aestivum* L.*)* cultivars. *Molecular Breeding* **23**:23-33

Charmet G (2011) Wheat domestication: lessons for the future. *C R Biol* **334**:212-220

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Paslier M-CL, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B and Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* **22**:1686-1701

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen H, Frazer KA, Huson DH, Schölkopf B, Nordborg M, Rätsch G, Ecker JR and Weigel D (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**:338-342

Clouse SD (1996) Plant hormones: brassinosteroids in the spotlight. *Curr Biol* **6**:658-661

Crawley MJ (2007) The R book. John Wiley & sons Ltd, England

Czembor PC, Arseniuk E, Czaplicki A, Song Q, Cregan PB and Ueng PP (2003) QTL mapping of partial resistance in winter wheat to Stagonospora nodorum blotch. *Genome* **46**:546-554

Devos KM (2005) Updating the 'crop circle'. *Curr Opin Plant Biol* **8**:155-162

Doležel J, Kubaláková M, Paux E, Bartoš J, Feullet C (2007) Chromosome-based genomics in the cereals. *Chromosome Res* **15**:51-66

Doležel J, Vrána J, Safář J, Bartoš J, Kubaláková M and Simková H (2012) Chromosomes in the flow to simplify genome analysis. *Funct Integr Genomics* **12**:397-416

Draeger R, Gosman N, Steed A, Chandler E, Thomsett M, Srinivasachary, Schondelmaier J, Buerstmayr H, Lemmens M, Schmolke M, Mesterhazy A and Nicholson P (2007) Identification of QTLs for resistance to Fusarium head blight, DON accumulation and associated traits in the winter wheat variety Arina. *Theor Appl Genet* **115**:617-625

Du CC, Nelson LR, McDaniel ME (1999) Diallel analysis of gene effects conditioning resistance to *Stagonospora nodorum* (Berk.) in wheat. *Crop Science* **39**:686-690

Duczek LJ, Sutherland KA, Reed SL, Bailey KL, Lanford GP (1999) Survival of leaf spot pathogens on crop residues of wheat and barley in Saskatchewan. *Can J Plant Pathol* **21**:165–173

Duveiller E, Singh RP, Nicol JM (2007) The challenges of maintaining wheat productivity: pests, diseases, and potential epidemics. *Euphytica* **157**:417-430

Eriksen L, Borum F and Jahoor A (2003) Inheritance and localisation of resistance to *Mycosphaerella graminicola* causing septoria tritici blotch and plant height in the wheat (*Triticum aestivum* L.) genome with DNA markers. *Theor Appl Genet* **107**:515-527

Eyal Z, Scharen AL, Prescott JM and van Ginkel M (1987) The *Septoria* diseases of wheat: concepts and methods of disease management. Mexico, D.F.:CIMMYT

FAO stat (2010) Food and Agriculture organization of the United Nations. Available at http://faostat.fao.org

Faris JD, Haen KM and Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot region in wheat. *Genetics* **154**:823-835

Faris JD, Zhang Z, Lu H, Lu S, Reddy L, Cloutier S, Fellers JP, Meinhardt SW, Rasmussen JB, Xu SS, Oliver RP, Simons KJ and Friesen TL (2010) A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proc Natl Acad Sci USA* **107**:13544-13549

Fernandez MGS, Bercraft PW, Yin Y, Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. *Trends Plant Sci* **14(8)**:454-461

Feuillet C, Stein N, Rossini L, Praud S, Mayer K, Schulman A, Eversole K and Appels R (2012) Integrating cereal genomics to support innovation in the *Triticeae*. *Funct Integr Genomics* **12**:573-583

Flor HH (1955) Host-parasite interaction in flax rust - its genetics and other implications. *Phytopathology* **45**:680-685

Francki MG, Shankar M, Walker E, Loughman R, Golzar H and Ohm H (2011) New quantitative trait loci in wheat for flag leaf resistance to Stagonospora nodorum blotch. *Phytopathology* **101**:1278-1284

Fridman E, Liu YS, Carmel-Goren L, Gur A, Shoresh M, Pleban T, Eshed Y and Zamir D (2002) Two tightly linked QTLs modify tomato sugar content via different

physiological pathways. *Mol Genet Genomics* **266**:821-826

Fried PM, Meister E (1987) Inheritance of leaf and head resistance of winter-wheat to *Septoria-nodorum* in a diallel cross. *Phytopathology* **77**:1371-1375

Friesen TL, Chu C-G, Liu ZH, Xu SS, Halley S, Faris JD (2009) Host-selective toxins produced by Stagonospora nodorum confer disease susceptibility in adult wheat plants under field conditions. *Theor Appl Genet* **118**:1489-1497

Friesen TL, Meinhardt SW and Faris JD (2007) The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. *Plant J* **51**:681-692

Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA and Oliver RP (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet* **38**:953-956

Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, Chen X, Sela H, Fahima T and Dubcovsky J (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science* **323**:1357-1360

Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh C-T, Wu W, Swanson-Wagner R, D'Ascenzo M, Millard T, Freeberg L, Aoyama N, Kitzman J, Burgess D, Richmond T, Albert TJ, Barbazuk WB, Jeddeloh JA and Schnable PS (2010) Repeat subtraction-mediated sequence capture from a complex genome. *Plant J* **62**:898-909

Fujioka S and Yokota T (2003) Biosynthesis and metabolism of brassinosteroids. *Annu Rev Plant Biol* **54**:137-164

Gale MD and Devos KM (1998) Comparative genetics in the grasses. *Proc Natl Acad Sci USA* **95**:1971-1974

Gendron JM and Wang Z-Y (2007) Multiple mechanisms modulate brassinosteroid signaling. *Curr Opin Plant Biol* **10**:436-441

Gervais L, Dedryver F, Morlais J-Y, Bodusseau V, Negre S, Bilous M, Groos C and Trottet M (2003) Mapping of quantitative trait loci for field resistance to Fusarium head blight in an European winter wheat. *Theor Appl Genet* **106**:961-970

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES and Nusbaum C (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**:182-189

Gu YQ, Coleman-Derr D, Kong X and Anderson OD (2004) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four *triticeae* genomes. *Plant Physiol* **135**:459-470

Halama P (2002) Mating relationships between isolates of *Phaeosphaeria nodorum* (anamorph *Stagonospora nodorum*) from geographical locations. *Eur J Plant*

Pathol **108**:593–596

Hallauer AR, Miranda Fo JB (1981) Quantitative genetics in maize breeding. The Iowa University Press, Ames, Iowa, USA

Hane JK, Lowe RGT, Solomon PS, Tan K-C, Schoch CL, Spatafora JW, Crous PW, Kodira C, Birren BW, Galagan JE, Torriani SFF, McDonald BA and Oliver RP (2007) Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *Plant Cell* **19**:3347-3368

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP and Stupar RM (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiol* **155**:645-655

Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J and Mayer KFX (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J* **69**:377-386

Hollister JD and Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**:1419-1428

Hong Z, Ueguchi-Tanaka M, Fujioka S, Takatsuto S, Yoshida S, Hasegawa Y, Ashikari M, Kitano H and Matsuoka M (2005) The Rice brassinosteroid-deficient dwarf2 mutant, defective in the rice homolog of *Arabidopsis* DIMINUTO/DWARF1, is rescued by the endogenously accumulated alternative bioactive brassinosteroid, dolichosterone. *Plant Cell* **17**:2243-2254

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R and Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci USA* **99**:8133-8138

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T and Han B (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res* **19**:1068-1076

Huang X, Lu T and Han B (2013) Resequencing rice genomes: an emerging new era of rice genomics. *Trends Genet* **29**:225-232

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F, Li J and Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* **42**:961-967

Ishimaru K, Ono K and Kashiwagi T (2004) Identification of a new gene controlling plant height in rice using the candidate-gene strategy. *Planta* **218**:388-395

Jeger MJ, Viljanen-Rollinson SLH (2001) The use of the area under the disease progress curve (AUDPC) to assess quantitative disease resistance in crop cultivars. *Theor Appl Genet* **102**:32–40

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KFX, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J, Consortium IWGS, Yang H, Liu X, He Z, Mao L and Wang J (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**:91-95

Johnson R. (1983) Genetic Background of Durable Resistance. In: Lamberti F, Waller J and Graaff N (Ed.), Durable Resistance in Crops. Springer New York

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Vries HB-D, Effgen S, Vreugdenhil D and Koornneef M (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* **175**:891-905

Kilian B, Ozkan H, Deusch O, Effgen S, Brandolini A, Kohl J, Martin W and Salamini F (2007) Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. *Mol Biol Evol* **24**:217-227

Kleijer G, Bronnimann A, Fossati A (1977) Chromosomal location of a dominant gene for resistance at seedling stage to *Septoria-nodorum* berk in wheat variety Atlas-66. *Zeitschrift Fur Pflanzenzuchtung – Journal of Plant Breeding* **78**:170-173

Kosambi D (1944) The estimation of map distances from recombination values. *Ann Eugen* **12***:172–175*

Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL and Keller B (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* **323**:1360-1363

Krishna P (2003) Brassinosteroid-Mediated Stress Responses. *J Plant Growth Regul* **22**:289-297

Kroymann J and Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**:95-98

Kubaláková M, Vrána J, Číhalíková J, Šimková H and Doležel J (2002) Flow karyotyping and chromosome sorting in bread wheat *(Triticum aestivum* L.*)*. *Theor Appl Genet* **104**:1362-1372

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, Jiao Y, Ni P, Zhang J, Li D, Guo X, Ye K, Jian M, Wang B, Zheng H, Liang H, Zhang X, Wang S, Chen S, Li J, Fu Y, Springer NM, Yang H, Wang J, Dai J, Schnable PS and Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines.

Nat Genet **42**:1027-1030

Lai K, Berkman PJ, Lorenc MT, Duran C, Smits L, Manoli S, Stiller J and Edwards D (2012) WheatGenome.info: an integrated database and portal for wheat genome information. *Plant Cell Physiol* **53(2)**:e2(1-7)

Lander EC, Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**:231-239

Laubscher FX, von Wechmar B and van Schalkwyk D (1966), Heritable Resistance of Wheat Varieties to Glume Blotch (*Septoria nodorum* Berk.). *Journal of Phytopathology* **56**:260–264

Lazniewska J, Macioszek VK, Kononowicz AK (2012) Plant-fungus interface: the role of surface structures in plant resistance ans susceptibility to pathogenic fungi. *Physiol Mol Plant P* **78**:24-30

Li C, Chen M, Chao S, Yu J, Bai G (2013) Identification of a novel gene, *H34*, in wheat using recombinant inbred lines and single nucleotide polymorphism markers. *Theor Appl Genet* DOI 10.1007/s00122-013-2118-5

Li J and Jin H (2007) Regulation of brassinosteroid signaling. *Trends Plant Sci* **12**:37–41

Lindhout P (2002) The perspectives of polygenic resistance in breeding for durable disease resistance. *Euphytica* **124**:217-226

Ling H-Q, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo M-C, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A and Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**:87-90

Liu Z, Zhang Z, Faris JD et al (2012) The cysteine rich necrotrophic effector SnTox1 produced by *Stagonospora nodorum* triggers susceptibility of wheat lines harboring *Snn1*. *PLoS Pathogens* **8(1)**:e1002467

Liu ZH, Faris JD, Meinhardt SW, Ali S, Rasmussen JB and Friesen TL (2004) Genetic and Physical Mapping of a Gene Conditioning Sensitivity in Wheat to a Partially Purified Host-Selective Toxin Produced by *Stagonospora nodorum*. *Phytopathology* **94**:1056-1060

Lorang JM, Sweat TA, Wolpert TJ (2007) Plant disease susceptibility conferred by a "resistance" gene. *Proc Natl Acad Sci USA* **104**:14861-14866

Lucas SJ, Akpınar BA, Kantar M, Weinstein Z, Aydınoglu F, Šafář J, Šimková H, Frenkel Z, Korol A, Magni F, Cattonaro F, Vautrin S, Bellec A, Bergès H, Doležel J and Budak H (2013) Physical Mapping Integrated with Syntenic Analysis to Characterize the Gene Space of the Long Arm of Wheat Chromosome 1A *PLoS ONE* **8(4)**:e59542

Ma H, Hughes GR (1995) Genetic-control and chromosomal location of *Triticum timopheevii*-derived resistance to *Septoria nodorum* blotch in durum-wheat. *Genome* **38**:332-338

Mago R, Spielmeyer W, Lawrence J, Lagudah S, Ellis G and Pryor A (2002) Identification and mapping of molecular markers linked to rust resistance genes located on chromosome 1RS of rye using wheat-rye translocation lines. *Theor Appl Genet* **104**:1317-1324

Mayer KFX, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R and Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* **23**:1249-1263

McDonald MC, Oliver RP, Friesen TL, Brunner PC and McDonald BA (2013) Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. *New Phytol* **199**:241-251

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasram B, Bustamante CD, Weigel D, Mackill, DJ, Bruskiewich RM, Rätsch G, Buell CR, Leung H and Leach JE (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci USA* **106**:12273-12278

Mengiste T (2012) Plant immunity to necrotrophs. *Annu Rev Phytopathol* **50**:267-294

Mester D, Ronin Y, Minkov D, Nevo E and Korol A (2003) Constructing large-scale genetic maps using an evolutionary strategy algorithm. *Genetics* **165**:2269-2282

Mesterhazy A (1995) Types and components of resistance to Fusarium head blight of wheat. *Plant Breeding* **114**: 377-386

Monna L, Lin X, Kojima S, Sasaki T and Yano M (2002) Genetic dissection of a genomic region for a quantitative trait locus, *Hd3*, into two loci, *Hd3a* and *Hd3b*, controlling heading date in rice. *Theor Appl Genet* **104**:772-778

Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS and Johal GS (2003) Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. *Science* **302**:81-84

Munkvold JD, Greene RA, Bermudez-Kandianis CE, Rota CML, Edwards H, Sorrells SF, Dake T, Benscher D, Kantety R, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorák J, Miftahudin, Gustafson JP, Pathan MS, Nguyen HT, Matthews DE, Chao S, Lazo GR, Hummel DD, Anderson OD, Anderson JA, Gonzalez-Hernandez JL, Peng JH, Lapitan N, Qi LL, Echalier B, Gill BS, Hossain KG, Kalavacharla V, Kianian SF, Sandhu D, Erayman M, Gill KS, McGuire PE, Qualset CO and Sorrells ME (2004)

Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. *Genetics* **168**:639-650

Murphy NEA, Loughman R, Wilson R et al (2000) Resistance to septoria nodorum blotch in the *Aegilops tauschii* accession RL5271 is controlled by a single gene. *Euphytica* **113**:227-233

Nagy ED, Lee T-C, Ramakrishna W, Xu Z, Klein PE, SanMiguel P, Cheng CP, Li J, Devos KM, Schertz K, Bennetzen JL (2007) Fine mapping of the *Pc* locus of *Sorghum bicolor*, a gene controlling the reaction to a fungal pathogen and its host-selective toxin. *Theor Appl Genet* **114**:961-970

Nelson LR, Gates CE (1982) Genetics of host plant resistance of wheat to *Septoria nodorum*. *Crop Sci* **22**:771–773

Newey LJ, Caten CE and Green JR (2007) Rapid adhesion of *Stagonospora nodorum* spores to a hydrophobic surface requires pre-formed cell surface glycoproteins. *Mycol Res* **111**:1255-1267

Oerke EC, Dehne HW, Schoenbeck F, Weber A (1994) Crop production and crop protection: Estimated losses in major food and cash crops. Elsevier Science Publishers, Amsterdam

Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ and Zwick ME (2007) Microarray-based genomic selection for high-throughput resequencing. *Nat Methods* **4**:907-909

Oliver RP and Solomon PS (2010) New developments in pathogenicity and virulence of necrotrophs. *Curr Opin Plant Biol* **13(4)**:415-419

Oliver RP, Friesen TL, Faris JD and Solomon PS (2012) *Stagonospora nodorum*: from pathology to genomics and host resistance. *Annu Rev Phytopathol* **50**:23-43

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N and Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res* **18**:2024-2033

Paillard S, Schnurbusch T, Winzeler M, Messmer M, Sourdille P, Abderhalden O, Keller B and Schachermayr G (2003) An integrative genetic linkage map of winter wheat *(Triticum aestivum* L.*)*. *Theor Appl Genet* **107**:1235-1242

Parlevliet JE (2002) Durability of resistance against fungal, bacterial and viral pathogens: present situation. *Euphytica* **124**:147-156

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Rahman M, Ware D, Westhoff P, Mayer KFX,

Messing J and Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**:551-556

Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant J-P, Sourdille P, Balfourier F, Paslier M-CL, Chauveau A, Cakir M, Gandon B and Feuillet C (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol J* **8**:196-210

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P and Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J* **48**:463-474

Paux E, Sourdille P, Mackay I and Feuillet C (2012) Sequence-based marker development in wheat: advances and applications to breeding. *Biotechnol Adv* **30**:1071-1088

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Bergès H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M and Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* **322**:101-104

Pea G, Paulstephenraj P, Canè MA, Sardaro MLS, Landi P, Morgante M, Porceddu E, Pè ME and Frascaroli E (2009) Recombinant near-isogenic lines: a resource for the mendelization of heterotic QTL in maize. *Mol Genet Genomics* **281**:447-457

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD and Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**:256-261

Peng JH, Sun D, Nevo E (2011) Domestication evolution, genetics and genomics in wheat. *Mol Breeding* **28**:281-301

Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM and Leamon JH (2006) Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* **7**:216-237

Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, Gao Y, Church GM and Shendure J (2007) Multiplex amplification of large sets of human exons. *Nat Methods* **4**:931-936

Quraishi UM, Murat F, Abrouk M, Pont C, Confolent C, Oury FX, Ward J, Boros D, Gebruers K, Delcour JA, Courtin CM, Bedo Z, Saulnier L, Guillon F, Balzergue S, Shewry PR, Feuillet C, Charmet G and Salse J (2011) Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat *(Triticum aestivum* L.*)*. *Funct Integr Genomics* **11**:71-83

R Development Core Team (2009). R: A language and environment for statistical

computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://R-project.org. ISBN 3-900051-07-0

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Beckert M and Charmet G (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat *(Triticum aestivum* L.*)*. *Genome* **49**:1131-1139

Rustenholz C, Choulet F, Laugier C, Šafář J, Šimková H, Doležel J, Magni F, Scalabrin S, Cattonaro F, Vautrin S, Bellec A, Bergès H, Feuillet C and Paux E (2011) A 3,000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. *Plant Physiol* **157**:1596-1608

Šafář J, Šimková H, Kubaláková M, Číhalíková J, Suchánková P, Bartoš J, Doležel J (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet Genome Res* **129:**211-223

Saintenac C, Falque M, Martin OC, Paux E, Feuillet C and Sourdille P (2009) Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat *(Triticum aestivum* L.*)*. *Genetics* **181:**393-403

Saintenac C, Jiang D and Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol* **12**:R88

Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R and Martin W (2002) Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet* **3**:429-441

Salmon A, Udall JA, Jeddeloh JA and Wendel J (2012) Targeted capture of homoeologous coding and noncoding sequence in polyploid cotton. *G3 (Bethesda)* **2**:921-930

Sarkar P and Stebbins GL (1956) Morphological evidence concerning the origin of the B genome in wheat. *Am J Bot* **43**: 297–304

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H and Matsuoka M (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* **416**:701-702

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L,

Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Buren PV, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA and Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**:1112-1115

Schnurbusch T, Paillard S, Fossati D, Messmer M, Schachermayr G, Winzeler M and Keller B (2003) Detection of QTLs for Stagonospora glume blotch resistance in Swiss winter wheat. *Theor Appl Genet* **107**:1226-1234

Scott PR, Benedikz PW, Cox CJ (1982) A genetic study of the relationship between height, time of ear emergence and resistance to *Septoria nodorum* in wheat. *Plant Pathol* **31**:45–60

Scott PR, Benedikz PW, Zones HG, Ford MA (1985) Some effects of canopy structure and microclimate on infection of tall and short wheats by *Septoria nodorum*. *Plant Pathol* **34**:587–593

Shah DA and Bergstrom GC (2002) A Rainfall-Based Model for Predicting the Regional Incidence of Wheat Seed Infection by *Stagonospora nodorum* in New York. *Phytopathology* **92**:511-518

Shankar M, Walker E, Golzar H, Loughman R, Wilson RE and Francki MG (2008) Quantitative trait loci for seedling and adult plant resistance to Stagonospora nodorum in wheat. *Phytopathology* **98**:886-893

Shatalina M, Wicker T, Buchmann JP, Oberhaensli S, Simková H, Doležel J and Keller B (2013) Genotype-specific SNP map based on whole chromosome 3B sequence information from wheat cultivars Arina and Forno. *Plant Biotechnol J* **11**:23-32

Shaw MW, Bearchell SJ, Fitt BDL and Fraaije BA (2008) Long-term relationships between environment and abundance in wheat of *Phaeosphaeria nodorum* and *Mycosphaerella graminicola*. *New Phytol* **177**:229-238

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton J-M, Rees DJG, Williams KP, Holt SH, Rojas JJR, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troggio M, Viola R, Ashman T-L, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant DW, Fox SE, Givan SA, Wilhelm LJ, Naithani S, Christoffels A, Salama DY, Carter J, Girona EL, Zdepski A, Wang W, Kerstetter RA, Schwab W, Korban SS, Davik J, Monfort A,

Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE and Folta KM (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* **43**:109-116

Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh PK, Singh S and Govindan V (2011) The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annu Rev Phytopathol* **49**:465-481

Somers DJ, Isaac P and Edwards K (2004) A high-density microsatellite consensus map for bread wheat *(Triticum aestivum* L.*)*. *Theor Appl Genet* **109**:1105-1114

Srinivasachary, Gosman N, Steed A, Hollins TW, Bayles R, Jennings P and Nicholson P (2009) Semi-dwarfing *Rht-B1* and *Rht-D1* loci of wheat differ significantly in their influence on resistance to Fusarium head blight. *Theor Appl Genet* **118**:695-702

Stein N, Herren G and Keller B (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. *Plant Breed* **120**:354–356

Tanabe S, Ashikari M, Fujioka S, Takatsuto S, Yoshida S, Yano M, Yoshimura A, Kitano H, Matsuoka M, Fujisawa Y, Kato H and Iwasaki Y (2005) A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, dwarf11, with reduced seed length. *Plant Cell* **17**:776-790

Teer JK and Mullikin JC (2010) Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet* **19**:R145-R151

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**:711-716

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**:763-768

The International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* **436**:793-800

Thomson MJ, Edwards JD, Septiningsih EM, Harrington SE and McCouch SR (2006) Substitution mapping of *dth1.1,* a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. *Genetics* **172**:2501-2514

Tommasini L, Schnurbusch T, Fossati D, Mascher F and Keller B (2007) Association mapping of Stagonospora nodorum blotch resistance in modern European winter wheat varieties. *Theor Appl Genet* **115**:697-708

Trick M, Adamski NM, Mugford SG, Jiang C-C, Febrer M and Uauy C (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol* **12**:14

Uphaus J, Walker E, Shankar M, Golzar H, Loughman R, Francki M and Ohm H (2007)

Quantitative Trait Loci Identified for Resistance to Stagonospora Glume Blotch in Wheat in the USA and Australia. *Crop Science* **47(5)**:1813-1822

Van Ginkel M, Rajaram S (1999) Breeding for resistance to the Septoria/Stagonospora blights of wheat. In: van Ginkel M, McNab A, Krupinsky JM (eds) Septoria and Stagonospora diseases of cereals: a compilation of global research. CIMMYT, Mexico

Vrána J, Kubaláková M, Šimková H, Číhalíková J, Lysák MA and Doležel J (2000) Flow sorting of mitotic chromosomes in common wheat *(Triticum aestivum* L.*)*. *Genetics* **156**:2033-2041

Wanjugi H, Coleman-Derr D, Huo N, Kianian SF, Luo M-C, Wu J, Anderson O and Gu YQ (2009) Rapid development of PCR-based genome-specific repetitive DNA junction markers in wheat. *Genome* **52**:576-587

Waters ODC, Lichtenzveig J, Rybak K et al (2011) Prevalence and importance of sensitivity to the *Stagonospora nodorum* necrotrophic effector SnTox3 in current Western Australian wheat cultivars. *Crop & Pasture Science* **62**:556-562

Wenzl P, Suchankova P, Carling J, Simkova H, Huttner E, Kubalakova M, Sourdille P, Paul E, Feuillet C, Kilian A and Dolezel J (2010) Isolated chromosomes as a new and efficient source of DArT markers for the saturation of genetic maps. *Theor Appl Genet* **121:**465–474

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Šimková H, Kubaláková M, Choulet F, Tauden S, Platzer M, Feullet C, Fahima T, Budak H, Dolezel J, Keller B and Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley and their relatives. *Plant Cell* **23**:1706-1718

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M and Stein N (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant J* **59**:712-722

Wicki W, Winzeler M, Schmid JE, Stamp P, Messmer M (1999) Inheritance of resistance to leaf and glume blotch caused by *Septoria nodorum* Berk. in winter wheat. *Theor Appl Genet* **99**:1265–1272

Wolf EDD and Francl LJ (2000) Neural network classification of tan spot and Stagonospora blotch infection periods in a wheat field environment. *Phytopathology* **90**:108-113

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J and Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol* **30**:105-111

Yahiaoui N, Srichumpa P, Dudler R and Keller B (2004) Genome analysis at different ploidy

levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. *Plant J* **37**:528-538

Yamamoto T, Nagasaki H, ichi Yonemaru J, Ebana K, Nakajima M, Shibaya T and Yano M (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics* **11**:267

Zellinger E, Hawes CR, Gurr SJ, Dewey FM (2006) Attachment and adhesion of conidia of *Stagonospora nodorum* to natural and artificial surfaces. *Physiol Mol Plant P* **68**:209-215

Zentella R, Zhang Z-L, Park M, Thomas SG, Endo A, Murase K, Fleet CM, Jikumaru Y, Nambara E, Kamiya Y and Sun T-P (2007) Global analysis of della direct targets in early gibberellin signaling in *Arabidopsis*. *Plant Cell* **19**:3037-3057

Zhang Z, Friesen TL, Simons KJ, Xu SS, Faris JD (2009) Development, identification, and validation of markers for marker-assisted selection against Stagonospora nodorum toxin sensitivity genes Tsn1 and Snn2 in wheat. *Mol Breeding* **23**:35-49

Zohary D, Hopf M. Domestication of Plants in the Old World – The origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley (2001) 3rd edition. Oxford University Press, Oxford

ACKNOWLEDGEMENTS

I would like to thank...

...Beat Keller for his mentorship and support, for the opportunity to work on a challenging project, for always keeping his door open for discussions, for allowing me to learn as much as I wish, and for the patiently correcting the“the”s and “a”s in all my texts.
...Thomas Wicker(feld) for excellent guidance through the world of wheat genomics, for having more confidence in my bioinformatic skills than I did, for his optimism, awesome designer collections, and for being part of my PhD committee.
...Bruce McDonald for the interesting discussions, for helping me to see the situation from the other (fungal) side and for being part of my PhD committee.
...Robert Dudler for the nice discussions at the Apéros, for being my co-examiner and taking interest in my project.
..all the members of Keller's, Ringli's and Dudler's groups for the nice creative atmosphere, fruitful discussions and technical help – thank you Gabi, Geri and Agnesza for running the lab smoothly.
...Bea for taking care of my plants, helping me with all field seasons and motivating me to speak German.
...members of TriticeaeGenome project: I enjoyed being a part of this friendly and supportive community. Thanks for many inspiring meetings, collaborations and helpful courses.
...my favorite book-chapter-angels Lisi and Jyoti for being a wonderful team and sharing all the ups and downs of writing.
...Javier for his enthusiasm and for taking over the project.
...Simon for guiding me in the beginning of my work and for the patient help with this Thesis.
...Stefan for translating the Summary and daily support of nearly-finishing-PhDs.
...all current and former inhabitants of P3/12: Simone, Jan, Jimmy, Kostas, Stefan, Chris, Fabrizio and Thomas for being the best friends and colleagues ever. Thanks for the inspiration, lame jokes, support, beer on the roof and for handling every problem or success as if it was their own. Special thanks goes to Jan for helping me with every computer/programming/Jan-it-doesn't-work problem as well as for his warm heart and great sense of humor. Another very special thanks goes to Simone for her constant support, sarcastic jokes, realistic view of life and free German lessons.
...Lisi, Francis, Simone, Kostas, Jan, Jimmy, Gabi, Matthias, Stefan, Elena and Jyoti for their friendship. Special thanks to “the ladies” Lisi, Simone, Elena and Jyoti for their creativity, many moments of joy, our girly chats and giggles.
...Opernhaus Zürich for the wonderful evenings and everybody who shared these moments with me.
...my dearest friends from Siberia – Masha, Maria and Dmitry for not caring about distances.
...my family (both its Russian and Swiss parts), especially my parents, Elena and Nikolay, and my sister Vera for their love and support.
...my husband Christian for always seeing a bright side and making me see it as well, for his constant encouragement and love.